UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS DE LA EDUCACIÓN
Departamento de Investigación y Psicología en Educación

TESIS DOCTORAL

El efecto de regresión a la media en evaluaciones longitudinales a través de un estudio de simulación Monte Carlo

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Delia Arroyo Resino

Directores

María Castro Morera
Enrique Navarro Asencio

Madrid
Ed. electrónica 2019

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE EDUCACIÓN

DEPARTAMENTO DE INVESTIGACIÓN Y PSICOLOGÍA EN EDUCACIÓN

EL EFECTO DE REGRESIÓN A LA MEDIA EN EVALUACIONES LONGITUDINALES A TRAVÉS DE UN ESTUDIO DE SIMULACIÓN MONTE CARLO

Memoria para optar al grado de doctor presentada por:

DELIA ARROYO RESINO

Bajo la dirección de:

Dra. María Castro Morera

Dr. Enrique Navarro Asencio

Madrid, 2018

Había una vez un joven que soñaba con ser un gran inventor. Estudiaba día y noche, estudió varios años y, finalmente, escribió en su diario personal: "Ya he estudiado bastante. Soy ya un "himbestigador", y demostraré mi gran valía".

Comenzó de inmediato hacer experimentos y llegó a inventar los agujeros del queso. Pero pronto supo que ya habían sido inventados.

Volvió a comenzar desde el principio. Estudiaba mañana y tarde, estudió muchos meses y, finalmente escribió en su diario: "Ya es suficiente. Ahora soy de verdad un "himbestigador". El mundo verá de lo que soy capaz".

Y, en efecto, el mundo pudo ver: inventó los agujeros del paraguas y fue el hazmerreir de todos.

Pero él no se desanimó. Volvió sobre los libros, rehízo experimentos tras experimentos y finalmente escribió en su diario:
"Bien, ahora estoy seguro de no equivocarme. Ahora soy un "himbestigador" en serio.

En cambio, era ahora un "himbestigador" con una pequeña falta. Inventó una nave que viajaba impulsada por pintura al pastel, costaba demasiado y coloreó todo el mar.
- No me detendré por ello- se dijo el buen joven, que ya comenzaba a tener canas.

Estudió, estudió y estudió tanto que llegó a ser un investigador con todas las letras en su puesto, y así pudo inventar todo lo que quiso. Inventó un vehículo para viajar a la Luna, un tren que sólo consumía un grano de arroz cada mil kilómetros, los zapatos que no se gastan nunca, y muchas cosas más.

Pero el sistema de llegar a ser investigadores y científicos sin cometer errores no llegó a inventarlo ni siquiera él, y tal vez no lo invente nadie nunca.
GIANNI RODARI (1964)

Agradecimiento

A medida que iba culminando la realización de la tesis doctoral, y pensando en lo que iba a escribir en dicho apartado, me daba cuenta de lo complejo que iba ser escribir estas líneas, en gran medida, por la cantidad de emociones que se despiertan en mí al recordar a todas aquellas personas que han estado a mi lado a lo largo de esta maravillosa y en ocasiones dificil experiencia, y sin las cuales el camino recorrido nunca hubiera sido igual de gratificante.

El comienzo de toda tesis doctoral es una tarea ardua llena de incertidumbre, pero el final tampoco es fácil, ya que supone poner fin a una etapa llena de aprendizajes, experiencias, emociones y conocimiento de uno mismo, que te hace reflexionar sobre aquellas personas que estuvieron ahí, ayudándote a luchar por tu meta. Sin esas personas, el trabajo que se presenta en las páginas siguientes nunca hubiera sido posible, por ello quiero mencionarlas aquí.

En primer lugar, de manera muy especial me gustaría dar gracias a mi directora la Dra. Maria Castro Morera, por haberme dado la oportunidad de trabajar con ella. Creo que una de las mejores decisiones que tomé en mi vida fue pedirte ayuda para la realización de mi primer proyecto fin de máster, ya que, gracias a ello, pude conocer a una persona maravillosa que en todo momento confio en mi y en la realización de dicha tesis doctoral, ella es todo un ejemplo a seguir a nivel profesional pero también a nivel personal.

En segundo lugar, también de manera muy especial quiero dar gracias a mi codirector el Dr. Enrique Navarro Asencio, he de decir que no conozco a hombre más paciente que él, las palabras de agradecimiento se me quedan cortas debido a todo el tiempo que me ha dedicado a resolver mis dudas, a guiarme en este trabajo de investigación, a hacer revisiones, en definitiva, a enseñarme a hacer una tesis doctoral.

Por supuesto tampoco puedo olvidar al Dr. José Luis Gaviria Soto, sin él esta tesis nunca hubiera sido posible, gracias por su generosidad a la hora de compartir conmigo su conocimiento tanto en sus clases como en reuniones informales, y gracias por promover en mí el espíritu investigador de continua búsqueda para la resolución de problemas.

No puedo dejar tampoco de dar las gracias a la Dra. Coral García Barbera y la Dra. Inmaculada Asencio Muñoz, a la primera de ellas me gustaría agradecerle el apoyo que me ha ofrecido. Desde el primer momento en que la conocí, siempre me ha tenido presente en su cabeza para mejorar mi carrera investigadora. A Inmaculada me gustaría agradecerle la escucha y el apoyo que da a los jóvenes investigadores para mejorar su formación, también agradecerla que siempre ha tenido la puerta de su despacho abierta para cualquier consulta, duda o propuesta, ¡echaré de menos ser vecinas de despacho!

Quiero dar gracias a todos los profesores del departamento de Investigación y Psicología en Educación de la Universidad Complutense de Madrid (antigo MIDE) especialmente a Covadonga Ruiz de Miguel, Ángeles Blanco Blanco, Celia Camilli Trujillo, Elvira Carpintero Molina, Chantal Becinto López y Luis Ayma González, por que siempre han estado ahí cuando he necesitado algo y han contado conmigo como profesora e investigadora. También quiero dar gracias al grupo de investigación MESE (Medida y Evaluación de Sistemas Educativos), la mayoría de ellos también forman parte del departamento (IPE). Gracias por permitirme trabajar con vosotros, ya que el principal problema de investigación de dicha tesis surge a partir de un trabajo que se estaba realizando con este grupo de investigación. Gracias también a Miguel Serra y a todo el personal de administración por su disponibilidad y estar siempre ahí ayudándonos en todo lo que hemos necesitado, y por preocuparse por todas nuestras necesidades. En este sentido también quiero agradecer a Ángeles Blanco Blanco como coordinadora de la línea de doctorado en la cual se inscribe esta tesis (Procesos sociales y evaluación de políticas educativas) por toda la ayuda recibida durante estos años, gracias por dar siempre una

respuesta eficiente y rápida a mis ruegos y preguntas, aunque en determinadas ocasiones solo fueran simples preocupaciones.

Tengo que acordarme también de todas mis compañeras de doctorado, especialmente de Judit, Silvia, Prado y Laura, por todas las charlas compartidas en las comidas, que muchas veces eran una fuente de liberación ante ciertos bloqueos, gracias por escucharme, apoyarme y aconsejarme. Por supuesto no puedo olvidar a mi compañero de despacho Víctor, con el cual he compartido una maravillosa etapa, gracias por alegrarme con tus canciones, por ayudarme cuando lo he necesitado, y, sobre todo, por tu carácter alegre, gracias por ser un gran compañero con el que contar.

Me gustaría dar las gracias a la Universidad KU Leuven (Bélgica) porque durante los tres meses que estuve allí me sentí como de la casa, especialmente quiero dar las gracias a mi tutor Wim Van den Noortgate, por permitirme trabajar con él.

A nivel más personal me gustaría agradecer en primer lugar a mis padres (Blanca y Eusebio) y hermanos (Jose y David) que hayan entendido la profesión que he elegido, sé que al principio fue difícil comprender que mi trabajo fuera siempre estudiar y que tuviera que viajar y salir de España, pero por fin saben lo importante que es para mí todo esto, muestra de ello es el apoyo y la ayuda continua que he recibido para llegar hasta aquí. En este sentido quiero hacer una especial mención a mi padre. Gracias papá por apoyarme y por estar siempre a mi lado en cada paso que he dado durante esta etapa, gracias por interesarte por lo que hacía, aunque no entendieras nada, gracias por venirte a vivir un tiempo conmigo a un país totalmente desconocido en el cual no dominabas su lengua, en definitiva, gracias por ser mi padre.

Gracias a mi pareja, Kelvin, por el apoyo que recibo día a día de él para que luche por mis sueños y que nunca tire la toalla, gracias por hacerme la vida más fácil, permitiendo que dedique más tiempo a mi trabajo, gracias por creer en mí, por compartir tu vida conmigo, espero que sigamos muchos años más cogidos de la mano.

Gracias a Lara y Alicia, por entender que muchas veces no pudiera hacer planes con vosotras porque tenía que terminar la tesis, sé que sois de las personas que más desean ver este proyecto culminado, junto con mi abuela, y de las que más apoyo me han dado tanto a nivel personal como profesional. Con vosotras, aunque estemos meses sin vernos cuando nos juntamos es como si el tiempo no hubiera pasado.

También me gustaría dar las gracias a mi pareja favorita Ara y Javi, porque siempre que os he necesitado habéis estado ahí, por escucharme, aconsejarme y ayudarme a tomar importantes decisiones en mi vida, gracias por ser tan buenas personas y tener tan buen corazón.

Quiero dar las gracias a mi amigo Sergio, conocerte fue una de las mejores cosas del máster de metodología, gracias por la paciencia que has tenido conmigo y por estar ahí siempre que te he necesitado, contigo realicé mi primera comunicación y eso nunca lo olvidaré.

Gracias a Marcos, porque desde el primer momento que le conté que iba a hacer una tesis doctoral, me apoyó en este gran reto, y me animó para que aprovechara todas las oportunidades que se me iban presentando en cada etapa de este proceso.

Finalmente, gracias también a Óscar y Yaiza por preocuparse por mí y entender y pasar conmigo mis "comeduras de cabeza", y siempre darme buenos consejos para calmar mi ansiedad.

Espero no olvidarme de nadie, pero si es así, aunque no os haya puesto aquí, tenéis mi agradecimiento.

¡MUCHAS GRACIAS!

ÍNDICE

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

LISTADO DE ABREVIATURAS UTILIZADO

AYP	Adequate Yearly Progress
CC	Calibración Conjunta
CF	Calibración Fija
CS	Calibración por Separado
CCIs	Curvas Características de los Ítems
ERM	Efecto de Regresión a la Media
EAP	Esperada a Posteriori
ESSA	Every Student Succeeds Act
ESO	Educación Secundaria Obligatoria
EVAAS	Evaluation Value Added Assessment System
HLM	Hierarchical Linear Models
IEA	International Association for the Evaluation of Educational Achievement
INEE	Instituto Nacional de Evaluación Educativa
LLECE	Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación
MAP	Máxima a Posteriori
MV	Máxima Verosimilitud
MLM	Mixed Linear Models
MVA	Modelo de Valor Añadido
ML4P	Modelo Logístico de cuatro Parámetros
ML2P	Modelo Logístico de dos Parámetros
ML3P	Modelo Logístico de tres Parámetros
ML1P	Modelo Logístico de un Parámetro
NCBL	No Child Left Behind
NWEA	Northwest Evaluation Association
OECD	Organization for Economic Cooperation and Development
SACMEQ	Southern and Eastern Africa Consortium for Monitoring Educational Quality
TAI	Test Adaptativo Informatizad
TCT	Teoría Clásica de los Test
TVAAS	Tennessee Value Added Assessment
TRI	Teoría de la Respuestas al Ítem
TVAAS	Tennessee Value-Added Assessment System
VA	Valor Añadido

RESUMEN

Las evaluaciones educativas cada vez están adquiriendo una mayor importancia como medio de control y mejora de la calidad de los sistemas educativos. A nivel metodológico existen diferentes tipos de evaluaciones, de estatus, basadas en un único momento temporal, evaluaciones de ganancia, caracterizadas porque se estudia el cambio en términos de rendimiento entre dos mediciones realizadas a un mismo estudiante, y evaluaciones basadas en modelos de crecimiento, que son las más adecuadas, pues el aprendizaje implica una evolución a lo largo del tiempo, y, por lo tanto, debería analizarse esa trayectoria académica de un mismo estudiante en diferentes ocasiones de medida.

A pesar de ser esta última evaluación la más idónea por la gran información que proporciona, es escaso el número de evaluaciones existentes de este tipo, debido a la complejidad y el coste que suponen, y menor es aún el número de estudios que analizan la fiabilidad existente en la estimación de las puntuaciones de los estudiantes obtenidos a lo largo del tiempo, así como la incidencia de posibles artefactos estadísticos. Uno de los destacados, propio de este tipo de diseños longitudinales, y que se aborda en este trabajo es el Efecto de Regresión a la Media.

En este sentido, en dicha tesis se pretende conocer cómo afectan determinadas características de las evaluaciones educativas, en los patrones de correlación generados entre las puntuaciones de los estudiantes, a través de la realización de una simulación Monte Carlo, basada en un modelo jerárquico lineal con tres niveles de anidamiento (tiempo, estudiante y aula) y longitudinal (nueve mediciones). Por ello, como **primer objetivo general** se plantea **estudiar el efecto de los porcentajes de varianza, así como de los valores de correlación fijados entre el intercepto y la pendiente en los distintos niveles, con el fin de conocer la fiabilidad con la que se estiman las puntuaciones de los estudiantes y el comportamiento del Efecto de Regresión a la Media.** Para dar respuesta a dicho objetivo se analizan los patrones de correlación generados entre las diferentes ocasiones de medida, lo que ayudará a conocer el número de aplicaciones necesarias para estimar las puntuaciones de los estudiantes con precisión.

Además, en este trabajo, se muestra una cierta preocupación por el tamaño muestral, ya que, en los diseños multinivel se recomienda un mínimo de 20 sujetos

por unidad de análisis, con el fin de que no exista mucho error en la estimación de la pendiente de crecimiento del nivel superior (generalmente aula o escuelas), algo que a veces es complicado debido a la mortandad experimental característica de los diseños longitudinales. Así, con el fin de conocer cómo el número de estudiantes por aula influye en la precisión con la que se estima la pendiente media de crecimiento, se plantea como **segundo objetivo general** de investigación, **estudiar el efecto del tamaño de las aulas en la estimación de su pendiente de crecimiento.**

Para la consecución de dichos objetivos de investigación, se han simulado datos con el software estadístico R-Studio versión 3.4.3, realizando más de 100 modelos multinivel, donde el primer nivel hace referencia al tiempo (nueve ocasiones de medida), el segundo a los estudiantes y, el tercero a las aulas, también se ha utilizado el programa estadístico SPSS versión 23 y Excel 2016, con el fin de complementar los análisis de los datos.

Los **resultados** muestran, en relación al **primer objetivo**, que el factor que más influye en la precisión con la que se estiman las puntuaciones de los estudiantes, es la diferencia entre el porcentaje de varianza encontrado en el error intrasujeto, respecto al porcentaje de varianza existente en la pendiente, obteniéndose estimaciones más fiables con un menor número de mediciones cuando es mayor este último coeficiente (debido a su carácter acumulativo), independientemente del nivel al que haga referencia (estudiante y/o aulas). Los valores de correlación existentes entre el intercepto y la pendiente no tienen un efecto muy marcado sobre la fiabilidad de las estimaciones en las puntuaciones de los estudiantes. Además, se ha observado que cuando se trabaja con datos propios de la realidad educativa, cuatro mediciones son suficientes para obtener estimaciones bastante fiables de las puntuaciones de los estudiantes (correlaciones superiores a 0,90). Respecto al Efecto de Regresión a la media, aunque dicho efecto desaparece a partir de la segunda ocasión de medida cuando se sigue a los mismos estudiantes en más de dos mediciones, dentro de la primera aplicación se observa, en la mayoría de los modelos simulados, que se mantiene constante o se incrementa a lo largo del tiempo.

Referido al **segundo objetivo**, los resultados revelan que las evaluaciones compuestas de un gran número de aulas formadas por más de 19 estudiantes, cuentan con un mayor porcentaje de aulas que se diferencian de la media, mientras

que, si se trabaja con aulas compuestas por menos de 20 estudiantes, la diferencia entre el porcentaje de aulas que si se diferencian de la media respecto al que no, es muy elevado, siendo notablemente más alto este último.

A la vista de dichos resultados se puede **concluir** que, generalmente, cuatro aplicaciones son suficientes para poder predecir el rendimiento de los estudiantes con precisión. En cuanto al Efecto de Regresión a la Media, las evaluaciones educativas donde se siguen a los mismos estudiantes en más de dos ocasiones de medida, los resultados finales están exentos de dicho fenómeno. Referido al tamaño muestral de las aulas, cuanto mayor sea el tamaño de las mismas, más precisas son las estimaciones de su pendiente de crecimiento. Por lo que se recomienda, no comparar el rendimiento de aulas con pocos estudiantes (menos de 20) con otro tipo de aulas compuestas por un mayor número de estudiantes, ya que, en la mayoría de las aulas formadas por pocos estudiantes, no es posible conocer su crecimiento real debido al error de medida. En definitiva, las evaluaciones educativas, por toda la complejidad que suponen, deben ser entendidas como una herramienta de ayuda que permitan potenciar las fortalezas y superar las dificultades de los sistemas educativos.

Palabras clave: evaluaciones educativas, evaluaciones longitudinales, modelos jerárquicos lineales, simulación Monte-Carlo

ABSTRACT

Educational assessments are acquiring increasing importance as a means of monitoring and improving the quality of education systems. On the methodological level, there are different types of assessment: status assessments, which are based on a single point in time; progress assessments, which study the change in terms of performance between two measurements applied to the same student; and assessments based on growth models, which are the most suitable, because if learning implies an evolution, then a student's academic progress should be analysed over a period which encompasses more than two measurements.

Despite the fact that this latter assessment is the most suitable due to the vast amount of information which it provides, very few assessments of this type exist due to their complexity and cost, and there are even fewer studies which analyse the reliability of estimates of student scores obtained over time or the impact of possible statistical artefacts. One of the most notable of these statistical artefacts is the Regression-to-the-Mean Effect, which is typical of this type of longitudinal design and is addressed in this work.

This thesis aims to discover how certain characteristics of educational assessments affect the correlation patterns generated between student scores, by means of a Monte Carlo simulation based on a linear (with three levels of nesting: time, student and class) and longitudinal (nine measurements) hierarchical model. Consequently, the **first general objective** of this research is **to study the effect of the percentages of variance and the fixed correlation values between the intercept and slope at the different nesting levels in order to discover how reliable estimates of student scores are and how the Regression-to-the-Mean effect behaves.** In order to achieve this objective, we analysed the correlation patterns generated between the different measurement occasions, which helps us to know the number of applications necessary in order to estimate student scores accurately.

Furthermore, this research work reflects a certain concern regarding sampling size, because, in multilevel designs, a minimum of 20 subjects per unit of analysis is recommended in order to eliminate as far as possible any errors in the

estimation of the growth curve of the upper level (usually class or school), and this can be complicated at times due to the experimental mortality which is characteristic of longitudinal designs. Thus, in order to know how the number of students per class affects the accuracy with which the mean growth curve is estimated, a **second general research objective** was proposed, which consisted of **studying how the size of a class affects the estimation of its growth curve.**

In order to achieve these research objectives, we simulated data using the statistical software R-Studio version 3.4.3 and carrying out more than 100 multilevel models where the first level refers to time (nine measurement occasions), the second to students, and the third to classes. We also used the statistical program SPSS version 23 and Excel 2016 to supplement the data analyses.

With regard to the **first objective**, the **results** shown that the factor which most affects the accuracy with which students' scores are estimated is the difference between the percentage of variance found in the intra-subject error and the percentage of variance in the slope, with more reliable estimates being obtained with a lower number of measurements when the latter coefficient is higher (due to its cumulative character), regardless of the level to which it refers (student and/or class). The correlation values between the variance of the intercept and the variance of the slope do not have a significant effect on the reliability of estimates of student scores. Furthermore, we observed that, when working with data obtained from real educational situations, four measurements are sufficient to obtain fairly reliable estimates of student scores (correlations greater than 0.90). As far as the Regression-to-the-Mean effect is concerned, although this effect disappears after the second measurement occasion when the same students are monitored over more than two measurements, in the majority of the simulated models we observe that, within the first application, the RME remains constant or increases over time.

With regard to the **second objective**, the results show that assessments made up of a large number of classes containing more than 19 students contain a higher percentage of classes that differ from the mean. However, with classes of fewer than 20 students, there is a significant difference between the percentage of classes which differ from the mean and the much larger percentage which do not.

In view of the said results, we can **conclude** that, generally, four applications are sufficient in order to be able to estimate student performance accurately. As for the Regression-to-the-Mean effect, this phenomenon does not occur in the final results of educational assessments where the same students are monitored over more than two measurement occasions. With regard to the sampling size of the classes, the larger the class, the more accurate the estimates of their growth curve. For this reason, we recommend that the performance of classes with few students (less than 20) not be compared with other types of class with a larger number of students, because, in the majority of classes containing few students, it is not possible to know their real growth due to measurement error. In short, educational assessments, given their complexity, must be understood as a tool which assists us in reinforcing the strengths and overcoming the difficulties of education systems.

Keywords: Educational assessments, longitudinal evaluations, hierarchical linear models, Monte Carlo simulation

.

INTRODUCCIÓN

Cada vez son más los países que utilizan las evaluaciones como un indicador de la calidad de los sistemas educativos. Pensar que es algo transitorio, carente de importancia, es un error, ya que, tal y como indicó la OECD en su publicación "*Education at a Glance, 2011: OECD indicators*", existe una tendencia al alza en el empleo de pruebas externas y estandarizadas (OECD, 2011).

En este sentido, a nivel nacional con la aparición de la actual ley de educación, Ley Orgánica de Mejora de la Calidad Educativa (LOMCE, 2013), se plantea la realización de una serie de evaluaciones externas o individualizadas, como medio de mejora y control de calidad del sistema educativo.

A nivel más internacional, con motivo del Foro Mundial sobre la Educación 2015, celebrado en Incheon (República de Corea), para el año 2030 se plantean una serie de objetivos educativos a través de los cuales se pretende proporcionar una educación de calidad, mediante procesos de evaluación longitudinales que permitan evaluar los progresos de los estudiantes a lo largo de tres ocasiones de medida (durante la enseñanza primaria, al final de la enseñanza primaria y al finalizar la enseñanza secundaria). Sin realmente conocer si tres medidas es un número lo suficientemente fiable como para proporcionar información rigurosa y de calidad, mediante la cual se puedan tomar decisiones educativas con cierto respaldo metodológico, ya que, no se han encontrado investigaciones sobre esta problemática.

Para garantizar la calidad de los sistemas educativos, en esta Declaración, y tal y como tradicionalmente ha venido ocurriendo en muchos estados norteamericanos, se han establecido marcos jurídicos y políticos que promueban la rendición de cuentas (accountability), y la transparencia, con el fin de informar a todos los agentes interesados en la educación, sobre el funcionamiento del sistema educativo, y así justificar la inversión del dinero público que se realiza con el mismo.

Generalmente la rendición de cuentas se ha entendido como "high-stakes" (alto impacto) o "low-stakes" (bajo impacto), en el primer caso, acorde a los resultados obtenidos en las evaluaciones educativas, las autoridades otorgaban premios o castigos a las escuelas. En el segundo caso, no se proporcionan este tipo de incentivos, sino que tienen una finalidad más informativa, es decir, mediante los

resultados de las evaluaciones se pretende únicamente conocer el estado del funcionamiento del sistema educativo.

Tanto en España, como inicialmente en la Declaración de Incheon, la rendición de cuentas es de "low-stakes", ya que, el objetivo es utilizar los resultados educativos como medio de mejora de la calidad de los sistemas de enseñanza. Por ello en dicha Declaración, la Unesco se compromete a desarrollar sistemas nacionales de seguimiento con evaluaciones integrales, que proporcionen datos lo suficientemente sólidos como para poder generar políticas y sistemas educativos eficientes e inclusivos, así como velar por la rendición de cuentas (UNESCO, 2016).

Con este fin, los países participantes deberán presentar informes al Instituto de Estadística de la Unesco como mecanismo de seguimiento del cuarto objetivo de desarrollo sostenible (ODS) (UNESCO, 2016), en el cual se establece "garantizar una educación inclusiva y equitativa de calidad y promover oportunidades de aprendizaje permanente para todos" (UNESCO, 2016, p.7).

El interés de este tipo de evaluaciones de *low-stakes* dista mucho del de las de *high-stakes*, que tienen su origen en las evaluaciones educativas desarrolladas en Estados Unidos con la promulgación en el año 2001 de la ley *No Child Left Behind* (NCBL). Con esta ley, la finalidad de dichas evaluaciones era sancionar a aquellas escuelas que no llegaban al progreso mínimo establecido (*adequate yearly progress*) y premiar a las que si. Fue abolida en el año 2015, cuando entró en vigor una nueva ley llamada Ley Cada Estudiante Triunfa (*Every Student Succeeds Act o ESSA*), que mantuvo ciertas partes de la NCLB, como era la rendición de cuentas y revocó otras, permitiendo a cada estado norteamericano establecer sus objetivos y evaluar los logros de las escuelas. Actualmente, con el mandato de Donald Trump se derrogan las normas que exigían a los centros y programas de entrenamiento de profesores, rendir cuentas al gobierno federal de los resultados obtenidos.

A pesar de la existencia de este tipo de evaluaciones, el objetivo de las mismas no debe ser sancionar o premiar a una escuela acorde a su rendimiento académico, sino debe ir encaminado a mejorar los sistemas y prácticas educativas, de tal manera, que se proporcione una educación de calidad basada en una evaluación continua, como parte elemental del proceso de enseñanza-aprendizaje (UNESCO, 2016). Este carácter de evaluación periódica, que permite analizar la evolución de

los agentes evaluados, busca la comparabilidad entre los resultados de las evaluaciones (Instituto de Evaluación, 2009), para lo que se requiere el establecimiento de ciertos criterios metodológicos que aseguren la fiabilidad en las estimaciones de las puntuaciones de los estudiantes, y una adecuada comparación de las mismas.

En este sentido, existen diversos procedimientos de comparación de puntuaciones a lo largo del tiempo, como es la construcción de escalas comunes, mediante métodos de equiparación y/o escalamiento. Generalmente el método utilizado para la comparación de puntuaciones en las distintas ocasiones de medida es el escalamiento vertical, ya que permite comparar los resultados de pruebas realizadas en distintos grados académicos, debido a que sitúa en una escala común los resultados de los cursos evaluados, de tal manera que es posible comparar las puntuaciones de los estudiantes a lo largo del tiempo, siempre que se mida el mismo constructo, aunque se cuente con una fiabilidad diferente.

Tradicionalmente las evaluaciones periódicas se han centrado en recoger información de los mismos cursos académicos, pero en años distintos, por lo que se analiza el rendimiento de cohortes diferentes en un único momento temporal. Se trata de evaluaciones de estatus que permiten evaluar el cambio que se produce de cohorte a cohorte, estudiando los cambios en las proporciones de estudiantes que se sitúan en los diferentes niveles de rendimiento académico en cada evaluación (Navarro, 2013).

Otra alternativa más eficiente y eficaz, es la que se propone en la Declaración de Incheon, que consiste en el seguimiento de una misma cohorte de estudiantes a lo largo del tiempo, algo poco común en muchos países, incluido España, donde este tipo de diseño fue realizado por primera vez por Marchesi, Martínez y Martín (2004) quienes realizaron un estudio longitudinal en el curso 1996-1997 de la información recogida con instrumentos ad-hoc, en una muestra de 31 escuelas de Educación Secundaria Obligatoria (ESO). En dicho estudio evaluaron el rendimiento de los estudiantes en Lengua, Matemáticas, Ciencias Sociales, Biología y Física y Química al inicio del primer curso de ESO y al final del segundo y cuarto curso, utilizando un modelo lineal mixto de medidas repetidas mediante el cual, se pretendía estudiar los cambios que se producían en los resultados académicos a lo largo del tiempo.

También Martínez-Arias, Gaviria y Castro (2009) realizaron una evaluación longitudinal en escuelas de la Comunidad de Madrid, durante los cursos 2005-2006 y 2006-2007 como parte de un proyecto de Investigación y Desarrollo [1], en el cual mediante un estudio piloto recogieron información de los mismos estudiantes al inicio y al final de dos cursos académicos en diferentes cohortes (5º y 6º de Educación Primaria, 1º y 2º de ESO y 3º y 4º de ESO) en las asignaturas de Comprensión Lectora y Matemáticas.

Aunque con dos medidas ya es posible conocer la ganancia que se produce en el rendimiento de los estudiantes entre una aplicación y otra (modelo de ganancia), es a partir de dos mediciones (modelo de crecimiento) cuando se puede definir la curva de crecimiento de los estudiantes (Raudenbush & Bryk, 2002). Pero hasta el momento no existe ningún acuerdo unánime que justifique el número de mediciones necesarias para la obtención de estimaciones fiables del rendimiento, teniendo en cuenta la presencia de ciertos artefactos estadísticos propios de los diseños longitudinales (por ejemplo, el efecto de regresión a la media). El Efecto de Regresión a la Media puede influir en las estimaciones de los niveles de logro de los estudiantes, pudiendo llegar a proporcionar resultados poco fiables o imprecisos sino se tiene en cuenta. Este fenómeno ha sido ampliamente estudiado, pero escasamente comprendido debido a la ambigüedad y a la gran cantidad de falacias que se han originado en torno al mismo, por ello como **principal objetivo teórico** de dicha tesis, se pretende **estudiar el Efecto de Regresión a la Media en los diseños longitudinales,** adaptando las explicaciones realizadas por Nesselroade, Stigler, y Baltes, (1980) en su artículo *Regression Toward the Mean and the Study of Change,* sobre el efecto de regresión a la media en un modelo de dos niveles (tiempo y estudiante) a un modelo de tres niveles (tiempo, estudiante y aula), mediante la utilización de ejemplos propios de la realidad educativa, que ayuden a una mejor comprensión de dicho fenómeno. Dicha explicación se considera una de las aportaciones teóricas más importantes de esta tesis, ya que, aunque el efecto de

[1] Son los resultados del proyecto I+D financiado por el Ministerio de Ciencia y Tecnología (Ref. SEC2003-09742) y titulado "El valor añadido en educación y la función de producción educativa: un estudio longitudinal". Su investigador principal fue el profesor José Luís Gaviria Soto y los desarrolló el grupo de investigación de la Universidad Complutense de Madrid MESE (Medida y Evaluación de Sistemas Educativos).

regresión es muy conocido, realmente no existen explicaciones sobre el mismo en un modelo de tres niveles.

Otros objetivos que se plantean a nivel teórico son:

a) Estudiar el crecimiento como una medida del aprendizaje
b) Definir el valor añadido y sus características metodológicas
c) Analizar el valor añadido en las evaluaciones educativas

Para el análisis metodológico del Efecto de Regresión a la Media, se ha optado por realizar una simulación experimental, que permita estudiar qué factores de los modelos de tres niveles de carácter longitudinal, pueden afectar y cómo afectan a la posible presencia de dicho artefacto estadístico. Sin una simulación sería imposible determinar en ausencia de un alto grado de incertidumbre, la presencia del Efecto de Regresión a la Media y los elementos de las grandes evaluaciones que favorecen su existencia, ya que, muchos de estos factores son imposible de manipular en la realidad educativa debido a que vienen impuestos por la misma.

En este sentido, como **primer problema de investigación** se plantea: **¿Cómo afecta la variabilidad de los coeficientes aleatorios y la correlación entre el intercepto y la pendiente cuando se trabaja con datos multinivel de carácter longitudinal, a la fiabilidad de las estimaciones de las puntuaciones de los estudiantes y al comportamiento del Efecto de Regresión a la Media?**

Este primer problema de investigación queda reflejado en un **primer objetivo general**, mediante el cual se pretende **simular una serie de modelos longitudinales y multinivel (modelos jerárquicos lineales) que permitan estudiar el efecto de los porcentajes de varianza, así como de los valores de correlación existentes entre el intercepto y la pendiente, sobre los patrones de correlación generados entre las distintas ocasiones de medida**. Dicho objetivo permitirá conocer, acorde con las características manipuladas en cada modelo, por un lado, el comportamiento del Efecto de Regresión a la Media en los diseños longitudinales en los cuales se sigue a los mismos sujetos en más de dos mediciones, y, por otro lado, la fiabilidad con la que se estiman las puntuaciones de los estudiantes, lo que ayudará a hacerse una idea del número de mediciones necesarias para poder obtener estimaciones fiables sobre el rendimiento real del estudiante.

Como **objetivos específicos** de este primer objetivo general de investigación, se plantean los siguientes:

1. Garantizar una adecuada recuperación de las condiciones experimentales fijadas, antes de la generación definitiva de los resultados, a través de una simulación Monte Carlo de un modelo jerárquico lineal con tres niveles de anidamiento (tiempo, estudiante y aula) y nueve ocasiones de medida.
2. Estudiar el efecto de los factores manipulados en los distintos modelos simulados, sobre la fiabilidad de las estimaciones de las puntuaciones de los estudiantes y sobre el comportamiento del Efecto de Regresión a la Media, bajo tres situaciones distintas:
    a) Cuando solo se fija varianza en algunos coeficientes aleatorios del modelo.
    b) Cuando además de fijar variabilidad en el error intrasujeto, en el intercepto y en la pendiente del nivel de aula o de estudiante, se fija también correlación entre el estatus y la pendiente de ese mismo nivel (uno de los dos niveles).
    c) Cuando además de fijar variabilidad en todos los coeficientes aleatorios de los tres niveles (tiempo, estudiante y aula), se fijan también distintos valores de correlación entre el intercepto y la pendiente del nivel de aula y del nivel de estudiante, de manera simultánea.

Además, como conclusión a este primer problema de investigación se pretende conocer, acorde a la relación existente entre el rendimiento de partida de las aulas y su pendiente de crecimiento, **las aulas que más potencian el rendimiento de los estudiantes (entre la primera y última ocasión de medida) agrupados en los diferentes quintiles**.

Este tipo de paradigma basado en el estudio del cambio a lo largo de diferentes ocasiones de medida, cuenta con una larga tradición que comienza con los trabajos de Rogosa y Willet (Rogosa & Willett, 1983; Willett, 1989).

Tanto los modelos de ganancia, basados en dos únicas mediciones (Demie, 2003; Jakubowski, 2008), como los de crecimiento, que utilizan más de dos aplicaciones (Castro, Ruíz & López, 2009; McCaffrey, Lockwood, Doretz & Hamilton,

2003a; Sanders & Horn, 1994; Singer & Willett, 2003; Stevens & Zvoch, 2006; Zvoch & Stevens, 2003), suponen un paso más allá de los modelos de estatus basados en un único momento temporal. Siendo considerados los modelos de crecimiento, por algunos autores (Garson, 2013; Singer & Willett, 2003; Stevens & Zvoch, 2006; Thum, 2009), como los más adecuados para evaluar el progreso de los estudiantes y la eficacia de las escuelas.

Los modelos de crecimiento son ideales para el estudio del Valor Añadido de las escuelas mediante el análisis de su residuo de crecimiento, ya que, facilitan el aislamiento de los efectos de las mismas o de los docentes, con el fin de comprobar su eficacia independientemente de otros factores ajenos al centro.

La metodología utilizada en los análisis del Valor Añadido tiene una larga tradición en EEUU, que también comenzó con el sistema de rendición de cuentas establecido en la ley NCBL. Asimismo el Estado de Tennesse (Sanders, Saxton & Horn, 1997), Dallas (Webster, 2005) o California (Doran & Izumi, 2004), entre otros, utilizan el Valor Añadido del docente o las escuelas como medio para analizar los datos obtenidos en las evaluaciones.

En Australia (Younk, 1999), Polonia (Jakubowski , 2008) o en España (Martínez-Arias, et al., 2009), se han realizado estudios de Valor Añadido con los resultados de las evaluaciones educativas, para informar y/o mejorar la calidad de los sistemas de enseñanza.

Uno de los mayores problemas de los modelos de valor añadido y/o crecimiento se refiere a la metodología utilizada en los mismos, ya que requiere de técnicas muy específicas que generalmente se ocultan en una "caja negra" (OECD, 2008, p.73) difícil de entender y, por lo tanto, ardua de implementar y/o explicar.

Concretamente la metodología utilizada en este tipo de modelos se centra en una modelización multinivel, donde generalmente se distinguen tres niveles de anidamiento: el tiempo, referido a las ocasiones de medida, el estudiante, ya que, en las distintas mediciones se suele evaluar el rendimiento de los estudiantes, y las aulas o escuelas que, es donde se encuentran agrupados los estudiantes. Estos modelos multinivel que maximizan el uso de la información disponible y minimiza algunos tipos de error, debido a que es posible obtener información muy precisa sobre el rendimiento del estudiante, se han venido estudiando desde la perspectiva

de los modelos jerárquico lineales (Raudenbush & Bryk, 2002), o bien, mediante modelos lineales mixtos con medidas repetidas (Sanders & Horn, 1994), ampliamente conocidos por ser el modelo base en la *Education Value-Added Assessment System* (EVAAS) de Tennessee.

Aunque los modelos multinivel o jerárquico lineales, en los que se basa el desarrollo metodológico de esta tesis, tal y como se ha puesto de manifiesto en el primer problema de investigación, son un caso particular de los modelos lineales mixtos (Rogosa & Willett, 1983), se diferencian principalmente en que mientras que en los modelos jerárquicos se representa la trayectoria individual de cambio, como una función de los parámetros de estatus inicial y crecimiento de un individuo, más el error aleatorio, con distribución normal, media cero y varianza común entre mediciones, en los modelos lineales mixtos, sin embargo, se estima la puntuación promedia en cada una de las aplicaciones, donde los términos residuales de las distintas puntuaciones de cada estudiante pueden estar correlacionados.

A pesar de las diferencias existentes entre dichas metodologías, ambos modelos forman parte de una misma moneda, a través de la cual se pretende estudiar el rendimiento de los estudiantes a lo largo del tiempo. Generalmente, en estos modelos multinivel el tamaño muestral juega un papel importante, ya que se recomienda un mínimo de entre 20 y 25 sujetos por unidad de análisis (Bryk & Raudenbush, 2002; Goldstein, 1997; Raudenbush & Bryk, 1986), algo que no siempre es fácil en la realidad educativa debido a la mortandad experimental propia de los diseños longitudinales. Teniendo en cuenta esta limitación, de manera más general, como **segundo problema de investigación** se plantea: **¿Afecta el tamaño de las aulas en la estimación de su pendiente de crecimiento?** Concretamente se pretende comprobar si es necesario tal y como apuntan los autores anteriores, un mínimo de al menos 20 estudiantes por aula para que las mismas se diferencien de la media, y, por lo tanto, se pueda conocer su crecimiento global.

Referido al segundo problema de investigación, cuyo interés reside en el estudio del tamaño de las aulas, se plantea como **objetivo general** del mismo, **a partir de los coeficientes de varianza y los valores de correlación fijados en uno de los modelos generados dentro del primer problema de investigación,**

estudiar el efecto del tamaño de las aulas en la estimación de la pendiente de crecimiento de las mismas.

En este segundo problema de investigación se plantean los siguientes **objetivos específicos**:

1. Garantizar una adecuada recuperación de las condiciones experimentales fijadas, antes de la generación definitiva de los resultados, a través de una simulación Monte Carlo de un modelo jerárquico lineal con tres niveles de anidamiento (tiempo, estudiante y aula) y nueve ocasiones de medida.
2. Estudiar el efecto del tamaño de las aulas sobre la precisión con la que se estima la pendiente de crecimiento de las mismas, en tres circunstancias distintas:
    a) Cuando de la muestra total de aulas, alrededor de un 75% son aulas pequeñas (comprendidas por un mínimo de 5 estudiantes y un máximo de 10 estudiantes), en torno a un 12,5% de las aulas son medianas (comprendidas por un mínimo de 11 estudiantes y un máximo de 19 estudiantes) y alrededor de otro 12,5% de las aulas son grandes (comprendidas por un mínimo de 20 estudiantes y un máximo de 30 estudiantes).
    b) Cuando la muestra total de aulas esta compuesta, en torno a un 75% por aulas medianas (comprendidas por un mínimo de 11 estudiantes y un máximo de 19 estudiantes), alrededor de un 12,5% por aulas pequeñas y alrededor de otro 12,5% por aulas grandes
    c) Cuando la muestra total de aulas esta compuesta, alrededor de un 75% por aulas grandes (comprendidas por un mínimo de 20 estudiantes y un máximo de 30 estudiantes), alrededor de un 12,5% por aulas pequeñas y en torno a otro 12,5% por aulas medianas

Para el cumplimiento de todos estos objetivos, y así, tratar de dar respuesta a los dos problemas de investigación planteados, que pueden ser de interés para la investigación basada en la realización y el estudio de las evaluaciones educativas, la tesis se estructura en **siete capítulos,** y **cuatro anexos** que debido a su extensión se presentan digitalizados en el pen drive que se adjunta.

Los cuatro primeros capítulos forman el cuerpo teórico del trabajo y describen aspectos relacionados con la preocupación educativa y social producida por las evaluaciones, la aparición y desarrollo de los análisis del Valor Añadido, las cuestiones metodológicas vinculadas a su medida y las diferentes aproximaciones que se utilizan para conseguir este propósito. Los tres siguientes capítulos, se dedican a los aspectos metodológicos de la tesis, donde se explica cómo son los procedimientos de simulación de los modelos generados, las condiciones de cada uno de los modelos, el análisis de resultados, discusión, conclusiones, limitaciones y prospectiva, entre otros aspectos.

Concretamente, el primer capítulo teórico, *Principales esquemas analíticos para la medida del crecimiento de los aprendizajes*, se centra en el estudio de la rendición de cuentas (accountability) y en los diferentes modelos utilizados para el análisis del cambio que se produce en los aprendizajes.

El segundo capítulo, como su propio título indica: *Aspectos metodológicos propios de los modelos de Valor Añadido*, se centra en el estudio de los aspectos metodológicos de dichos modelos, como son: la estructura de los datos, los procedimientos de comparación de puntuaciones, así como la importancia de las variables contextuales en este tipo de modelos, para predecir el rendimiento académico de los estudiantes.

En cuanto al tercer capítulo, *Estimación del valor añadido en las grandes evaluaciones educativas*, se continúa con el estudio del Valor Añadido, pero más centrado en los diferentes modelos utilizados para la estimación del mismo.

Finalmente, el último capítulo teórico de esta tesis es el cuarto, *La fiabilidad en las estimaciones de las puntuaciones de los estudiantes en los modelos longitudinales,* que en cierta medida es una introducción a la parte empírica, ya que se estudia fundamentalmente el Efecto de Regresión a la Media y las implicaciones que el mismo puede tener en la fiabilidad de las estimaciones de los niveles de logro de los estudiantes.

En cuanto al primer capítulo metodológico, capítulo cinco: *Diseño de una simulación Monte Carlo basada en modelos multinivel de carácter longitudinal,* se establecen los problemas y objetivos de investigación, así como los procedimientos desarrollados para dar respuesta a cada uno de los objetivos planteados.

En el capítulo seis, *Presentación de Resultados*, se analizan los resultados obtenidos en cada uno de los objetivos específicos, planteados dentro de los dos problemas de investigación.

Respecto al último capítulo, el siete: *Conclusiones, discusión, limitaciones y prospectiva*, se describen las conclusiones a modo de discusión, extraídas a lo largo de todo el trabajo. Además, se dan una serie de recomendaciones para la mejora de las evaluaciones educativas y se comentan algunas de las posibles limitaciones de la tesis y su prospectiva futura.

A partir de todo el contenido teórico y metodológico expuesto en esta tesis, se considera que la presente aportará a los especialistas e interesados en los sistemas educativos, especialmente en evaluaciones educativas de calidad, así como a las personalidades encargadas de las políticas educativas, adquirir un conocimiento más claro y práctico del comportamiento del Efecto de Regresión a la Media en las evaluaciones longitudinales, obtener información sobre cómo los factores de variabilidad y de correlación afectan en la precisión de las estimaciones de los niveles de logro de los estudiantes, y por ende, ayudar a conocer el número de mediciones más adecuado acorde a ciertas características de las evaluaciones educativas, así como ser consciente del efecto del tamaño muestral sobre el error de estimación de la pendiente de crecimiento de las aulas.

Se espera y se desea que las aportaciones de dicha tesis sirvan para el desarrollo de evaluaciones de calidad que mejoren el funcionamiento del sistema educativo en general. Sobre todo, en lo referente a la estimación de los niveles de logro de los estudiantes, con el fin de, tal y como se establece en la Declaración de Incheon, "proporcionar una educación inclusiva, equitativa y de calidad y un aprendizaje a lo largo de la vida para todos" (UNESCO, 2016, p.3)

INTRODUCTION

More and more countries are using **assessments** as an indicator of the quality of education systems. It would be a mistake to think that this is something transitory and lacking in important because, as the OECD indicated in its publication "*Education at a Glance, 2011: OECD indicators*", there is an upward trend in the use of **external and standardized tests** (OECD, 2011).

At the national level, the Education Act currently in force in Spain, the Organic Law for the Improvement of Educational Quality (LOMCE, 2013), provides for a series of external or individualized assessments as a means of improving and monitoring the quality of the education system.

On a more international level, the World Education Forum 2015, held in Incheon (Republic of Korea), established a series of educational objectives for the year 2030 with the intention of providing quality education by means of longitudinal evaluation processes which permit student progress to be assessed over three measurement occasions (during primary education, at the end of primary education and at the end of secondary education). However, we do not really know if three measurements are sufficient when it comes to providing comprehensive quality information which will serve as a methodological basis for educational decisions, as no research has been found which addresses this problem.

To ensure the quality of education systems, this Declaration, as is traditionally the case in many US states, established legal and political frameworks aimed at promoting accountability and transparency, in order to inform all the parties interested in education about the functioning of the education system, and thereby justify the corresponding investment of public money.

Generally, accountability has been understood as "high-stakes" or "low-stakes". In the first case, the authorities granted rewards to, or imposed punishments on, schools on the basis of the results obtained in educational assessments. In the second case, these types of incentive are not implemented as the assessments serve a more informative purpose, i.e. the results of the assessments are intended only to provide a picture of how the education system is functioning.

Both in Spain and, initially, in the Incheon Declaration, accountability is of a low-stakes variety since the objective is to use educational results as a means of improving the quality of teaching systems. Thus, the said Declaration expressed UNESCO's commitment to developing national monitoring systems with comprehensive assessments which provide solid enough data to generate efficient and inclusive education policies and systems as well as ensuring accountability (UNESCO, 2016).

To this end, participating countries are under an obligation to submit reports to the UNESCO Institute for Statistics as a mechanism for monitoring the fourth sustainable development goal (SDG) (UNESCO, 2016), which seeks to "ensure inclusive and equitable quality education and promote lifelong learning opportunities for all" (UNESCO, 2016, p. 7).

The objective pursued by this type of low-stakes assessment differs greatly from their high-stakes counterparts, which have their origin in the educational assessments drawn up in the United States with the passing in 2001 of the No Child Left Behind Act (NCLB). Under this law, the purpose of the said assessments was to penalize those schools which did not achieve adequate yearly progress and reward those which did. The law was abolished in the year 2015 and was replaced by the Every Student Succeeds Act (ESSA). This law retained certain parts of the NCLB, such as accountability, and revoked others, instead allowing each US state to establish its objectives and assess the achievements of its schools. Currently, under Donald Trump's mandate, the rules which required centres and teacher training programs to be accountable to the Federal Government for the results they obtained have been repealed.

In spite of the existence of this type of assessment, their objective should not be to penalize or reward schools on the basis of their academic performance but should be aimed at improving education systems and practices in such a way as to provide a quality education based on continuous assessment as a fundamental part of the teaching-learning process (UNESCO, 2016). This periodic assessment approach, which permits an analysis of the evolution of the agents assessed, seeks comparability between the results of the assessments (Institute for Evaluation, 2009), and this requires certain methodological criteria to be established in order

to ensure that estimations of student scores are reliable and that these scores can be suitably compared.

There are a range of procedures for comparing scores over time, such as the construction of common scales, using equivalence and/or scaling methods. Generally, vertical scaling is the method used to compare scores on different measurement occasions. This method permits a comparison of the results of tests carried out at different academic levels, because it situates the results of the courses assessed on a common scale, in such a way that it is possible to compare student scores over time, provided that the same construct is measured, even if their degree of reliability differs.

Traditionally, periodic assessments have focussed on compiling information about the same academic courses but in different years, and thus analyse the performance of different cohorts at a single point in time. These are status assessments which allow us to evaluate the change which occurs from cohort to cohort and study changes in the proportion of students who are situated at different levels of academic performance in each assessment (Navarro, 2013).

Another more efficient and effective alternative is the one proposed in the Incheon Declaration, which consists of monitoring the same cohort of students over time. This is not a very common practice in many countries, including Spain, where this type of design was implemented for the first time by Marchesi, Martínez and Martín (2004) when they carried out, in the academic year 1996-1997, a longitudinal study of the information compiled with ad hoc instruments in a sample of 31 schools providing compulsory secondary education (ESO). In this study, they assessed student performance in Language, Mathematics, Social Sciences, Biology, Physics and Chemistry at the beginning of the first year of ESO and at the end of the second and fourth year, using a mixed linear model of repeated measurements in order to study the changes that occurred in academic results over time.

Martínez-Arias, Gaviria and Castro (2009) also carried out a longitudinal evaluation in schools in the Madrid Region during the academic years 2005-2006 and 2006-2007 as part of a Research and Development project [2] in which they used

[2] These are the results of the R&D project financed by the Ministry of Science and Technology (REF. SEC2003-09742) and entitled "Value-added in education and the role of educational production: a

a pilot study to collect information from the same students at the beginning and end of two academic years in different cohorts (5th and 6th year of primary education, 1st and 2nd year of ESO and 3rd and 4th year of ESO) in the subjects of Reading Comprehension and Mathematics.

Although two measurements are sufficient in order to establish the progress that occurs in student performance between one application and another (progress model), the students' growth curve can only be defined with more than two measurements (growth model) (Raudenbush & Bryk, 2002). However, to date there is no unanimous agreement regarding the number of measurements necessary to obtain reliable performance estimates, bearing in mind the presence of certain statistical artefacts inherent to longitudinal designs (for example, the Regression-to-the-Mean Effect). The Regression-to-the-Mean Effect can influence estimations of the levels of student achievement and can even provide unreliable or imprecise results if it is not taken into account. This phenomenon has been widely studied but is scarcely understood due to the ambiguity and the large number of fallacies which have sprung up around it. For this reason, one **main theoretical objective** of this thesis is to **study the Regression-to-the-Mean Effect in longitudinal designs** by adapting the explanations provided by Nesselroade, Stigler and Baltes (1980) in their article *Regression Toward the Mean and the Study of Change* about the regression-to-the-mean effect in a two-level model (time and student) to a three-level model (time, student and class) and using examples from real situations in education to give us a better understanding of this phenomenon. We consider this to be one of the most important theoretical contributions of this thesis because, although the regression effect is well known, in reality there are no explanations about it in a three-level model.

Other objectives proposed at the theoretical level are:

a) To study growth as a measure of learning
b) To define value-added and its methodological characteristics
c) To analyse value-added in educational assessments

longitudinal study". The main researcher was Professor José Luís Gaviria Soto and the survey was carried out by the research group of the Complutense University of Madrid MESE (Measurement and Assessment of Educational Systems).

In order to conduct a methodological analysis of the Regression-to-the-Mean Effect, the decision was taken to carry out an experimental simulation which allows us to study which factors of the three-level longitudinal models can affect the possible presence of the said statistical artefact and how they affect it. Without a simulation it would be impossible to determine with any degree of certainty the presence of the Regression-to-the-Mean Effect and the elements of the large assessments which favour its existence because many of these factors are impossible to manipulate in real situations in education due to the fact that they are imposed by those situations.

The **first research problem** to arise is the following: **How do the variability of the random coefficients and the correlation between the intercept and slope affect the reliability of estimates of student scores and the behaviour of the Regression-to-the-Mean Effect when working with multilevel longitudinal data?**

This first research problem is reflected in a **first general objective**, which aims to **simulate a series of longitudinal multilevel models (hierarchical linear models) which will allow us to study the effect of the percentages of variance, and the covariance between the intercept and slope, on the correlation patterns generated between the different measurement occasions**. Based on the characteristics manipulated in each model, this objective will enable us to find out, on the one hand, the behaviour of the Regression-to-the-Mean Effect in the longitudinal designs in which the same subjects are monitored over more than two measurements, and, on the other hand , the reliability with which student scores are estimated, and this will help to give us an idea of the number of measurements needed in order to obtain reliable estimates of actual student performance.

The **specific objectives** of this first general research objective are as follows:

1. To ensure, prior to the final generation of results, an adequate recovery of the fixed experimental conditions by means of a Monte Carlo simulation of a hierarchical linear model with three levels of nesting (time, student and class size) and nine measurement occasions.

2. To study the effect of the factors manipulated in the various simulated models on the reliability of the estimates of student scores and on the behaviour of the Regression-to-the-Mean effect in three different situations:
   a) When variance is only fixed in a certain random coefficients of the model.
   b) When, in addition to fixing variance in the intra-subject error and in the intercept and slope of the class or student level, the correlation between the status and slope of that same level (one of the two levels) is also fixed.
   c) When, in addition to fixing variability in all the random coefficients of the three levels (time, student and class size), the various correlation values between the intercept and slope of the class level and the student level are fixed at the same time.

Furthermore, by way of conclusion to this first research problem, we aim to find out, based on the relationship between the initial performance of the classes and their growth curve, **the characteristics of the classes which most foster student performance (between the first and last measurement occasion) grouped in different quintiles**.

This type of paradigm, based on the study of change over a number of measurement occasions, enjoys a long tradition which began with the works of Rogosa and Willis (Rogosa & Willet, 1983; Willet, 1989).

Both the progress models, based on two single measurements (Demie, 2003; Jakubowski, 2008), and the growth models, which use more than two applications (Castro, Ruíz & López, 2009; McCaffrey, Lockwood, Doretz & Hamilton, 2003a; Sanders & Horn, 1994; Singer & Willet, 2003; Stevens & Zvoch, 2006; Zvoch & Stevens, 2003), constitute a step beyond the status models based on a single point in time. It is the growth models which are considered by some authors (Garson, 2013; Singer & Willet, 2003; Stevens & Zvoch, 2006; Thum, 2009) to be the most suitable for assessing student progress and school effectiveness.

Growth models are ideal for studying schools' value-added by analysing their residual growth, because they make it easier to isolate the effects of the schools themselves and of the teachers in order to verify their effectiveness independently of factors unrelated to the centre.

The methodology used in the analysis of value-added enjoys a long tradition in the US, and also began with the system of accountability established by the NCLB Act. Furthermore, the State of Tennessee (Sanders, Saxton & Horn, 1997), Dallas (Webster, 2005) and California (Doran & Izumi, 2004), among others, use the value-added of the teacher or the school as a means of analysing the data obtained in the assessments.

In Australia (Younk, 1999), Poland (Jakubowski, 2008) and Spain (Martínez-Arias, et al., 2009), value-added studies have been carried out with the results of educational assessments, in order to inform and/or steer education systems and policies towards educational quality.

One of the biggest problems of value-added and/or growth models arises from the methodology used therein, because it requires very specific techniques which are generally hidden in a "black box" (OECD, 2008, p. 73) which is difficult to understand and, therefore, arduous to implement and/or explain.

Specifically, the methodology used in these types of model focusses on a multilevel modelling approach, which generally differentiates between three levels of nesting: time, referring to the measurement occasions; the student, given that the various measurements usually assess student performance; and the classes or schools which is where students are grouped. These multilevel models, which maximize the use of the available information and minimize certain types of error, due to the fact that it is possible to obtain very accurate information about student performance, have been studied from the perspective of hierarchical linear models (Raudenbush & Bryk, 2002) or by means of mixed linear models with repeated measurements (Sanders & Horn, 1994), these being widely known as they constitute the base model in Tennessee's *Education Value-Added Assessment System* (EVAAS).

Although the multilevel or hierarchical linear models, which form the basis for the methodological development of this thesis, are, as highlighted in the first research problem, a particular case of mixed linear models (Rogosa & Willett, 1983),

they differ mainly in that whereas in the hierarchical models the individual trajectory of change is represented as a function of the parameters of initial status and growth of an individual, plus random error, with normal distribution, zero mean and common variance between measurements, in mixed linear model the average score is estimated in each of the applications, where the residual terms of the various scores of each student can be correlated.

Despite the differences which exist between these methodologies, the models form two sides of the same coin, which aims to study student performance over time. In these multilevel models, the sampling size generally plays an important role, as a minimum of between 20 and 25 subjects per unit of analysis is recommended (Bryk & Raudenbush, 2002; Goldstein, 1997; Raudenbush & Bryk, 1986). This is not always easy in real educational situations due to the experimental mortality characteristic of longitudinal designs. With this limitation in mind, and in a more general way, the **second research problem** is as follows: **Does the size of the class affect the estimation of its growth curve?** More specifically, the aim is to confirm whether, as indicated by the above authors, it is necessary to have at least 20 students per class so that the classes differ from the mean and we can therefore find out their overall growth.

The **general objective** the second research problem **is to study the effect of class size on the estimation of its growth curve, based on the fixed coefficients of variance and the correlation values in one of the models generated in the context of the first research problem.**

The **specific objectives** of this second research problem are as follows:

1. To ensure, prior to the final generation of results, an adequate recovery of the fixed experimental conditions by means of a Monte Carlo simulation of a hierarchical linear model with three levels of nesting (time, student and class size) and nine measurement occasions.
2. To study the effect of class size on the accuracy with which the growth curve is estimated in three different circumstances:
   a) When about 75% of the total sample of classes are small (made up of a minimum of 5 students and a maximum of 10), about 12.5% of the classes are medium-sized (made up of a minimum of 11 students and a maximum

of 19) and approximately another 12.5% of the classes are large (made up of a minimum of 20 students and a maximum of 30).

b) When around 75% of the total sample of classes consist of medium-sized classes (made up of a minimum of 11 students and a maximum of 19), about 12.5% are small classes and about another 12.5% are large classes.

c) When around 75% of the total sample of classes consists of large classes (made up of minimum of 20 students and a maximum of 30), about 12.5% are small classes and around another 12.5% are medium-sized classes.

In order to achieve these objectives, and thereby attempt to address the two research problems posed, which may be of interest to educational researchers charged with implementing and studying educational assessments, the thesis is structured in **seven chapters** and **four appendices** which, due to their length, are presented in digital format.

The first four chapters form the theoretical body of the work and describe aspects related to the educational and social concern generated by assessments, the emergence and development of value-added analyses, the methodological questions associated with their measurement and the different approaches applied to achieve this purpose. The following three chapters are devoted to the methodological aspects of the thesis, and explain the simulation procedures of the models generated, the conditions of each of the models, the analysis of results, discussion, conclusions, limitations, etc.

Specifically, the first theoretical chapter, *Main analytical frameworks for measuring learning growth*, focusses on the study of accountability and of the various models used to analyse the change which occurs in learning.

The second chapter, as indicated by its title, *Methodological aspects inherent to Value-Added Models*, focusses on the study of the more methodological aspects of these models, such as the structure of the data, the procedures for comparing scores, and the importance of contextual variables in this type of models when it comes to estimating students' academic performance.

In the third chapter, *Estimation of value-added in large educational assessments*, we continue with the study of value-added, but with a greater focus on the different models used for estimating it.

Finally, the fourth and last theoretical chapter of this thesis addresses *Reliability in the estimation of students' scores in longitudinal models* and is, to a certain extent, an introduction to the empirical part, as it deals mainly with the Regression-to-the-Mean Effect and the implications this can have for the reliability of estimates of the levels of student achievement.

The first methodological chapter, Chapter Five: *Design of a Monte Carlo simulation based on longitudinal multilevel models,* sets out the problems and objectives of the research, as well as the procedures applied to address each of the objectives established.

Chapter Six, *Presentation of Results*, analyses the results obtained in each of the specific objectives established as part of the two research problems.

The final chapter, Chapter Seven: *Conclusions, discussion, limitations and future prospects*, describes, in a discussion format, the conclusions reached over the course the work. In addition, it provides a number of recommendations for the improvement of educational assessments and comments on some of the possible limitations of the thesis and its future prospects.

On the basis of all the theoretical and methodological content presented, the author believes that the thesis will provide specialists and others who are interested in education systems with the following:

- A greater understanding of the behaviour of the Regression-to-the-Mean Effect in longitudinal evaluations;
- Information on how the factors of variability and correlation affect the accuracy of estimates of the levels of student achievement;
- Knowledge of the most suitable number of measurements on the basis of certain characteristics of educational assessments;
- Awareness of the effect of sampling size on the estimation error of the growth curve of the classes.

It is hoped that the contributions made by this thesis will serve to create quality assessments which improve the functioning of the education system in general and, above all, with regard to the estimation of levels of student achievement, and as set out in the Incheon declaration, to "ensure inclusive and equitable quality education and promote lifelong learning opportunities for all "(UNESCO, 2016, p.3).

PARTE TEÓRICA

CAPÍTULO I: PRINCIPALES ESQUEMAS ANALÍTICOS PARA LA MEDIDA DEL CRECIMIENTO DE LOS APRENDIZAJES

La evaluación de la educación es una herramienta fundamental para conocer el estado actual de los sistemas educativos, así como para la introducción de mejoras. Incluso los no partidarios de los sistemas de rendición de cuentas consideran que la evaluación es un componente necesario para el desarrollo de un buen sistema educativo.

Son numerosos los organismos nacionales e internacionales (*Instituto Nacional de Evaluación Educativa* (INEE), *International Association for the Evaluation of Educational Achievement* (IEA), *Organization for Economic Cooperation and Development* (OECD), *Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación* (LLECE), *Southern and Eastern Africa Consortium for Monitoring Educational Quality* (SACMEQ)) preocupados por el análisis académico de los estudiantes, en términos de rendimiento o bien en términos de competencia o en ambas, a pesar de que no hay una definición única ni unánime sobre los mismos, lo que hace que sean términos multidimensionales, contextuales y relativos (Salvador, Rodríguez & Bolivar, 2004) existiendo una gran variedad de definiciones propuestas por diversos autores apoyándose en distintas perspectivas.

Las repercusiones políticas y mediáticas de dichas evaluaciones de rendimiento y/o competencia se incrementan año tras año, así como su magnitud y diversidad. Lo que ha llevado al desarrollo de numerosas evaluaciones educativas que se realizan de forma periódica (PISA, TIMSS, PIAAC, etc.) con la finalidad de comprobar el funcionamiento de diferentes aspectos del sistema educativo y en cierta medida "rendir cuenta" a los organismos políticos. Esta idea de "rendir cuenta" o "*accountability*", tiene su origen en la aparición de leyes educativas, como la inglesa y norteamericana *No Child Left Behind*, mediante la cual se desarrolló un sistema de rendición de cuentas donde aquellas escuelas que no tenían una evaluación satisfactoria, por no conseguir los estándares prefijados, eran sancionadas y sometidas a acciones que mejoraran el nivel de logro de los centros educativos.

En este contexto los modelos de valor añadido, apoyados en el análisis de la ganancia o de crecimiento para la estimación de los resultados de los centros, aulas o docentes, han adquirido una gran importancia, ya que a través de los mismos es posible estudiar el efecto que las escuelas tiene sobre el nivel de logro de los estudiantes, aislando el resto de factores que pueden intervenir en los procesos educativos.

El valor añadido es considerado como la contribución neta en el progreso de los estudiantes hacia la consecución de unos objetivos de aprendizaje que han sido previamente definidos. Se habla de contribución neta porque se elimina la influencia de ciertos factores que, aunque se encuentren fuera del control escolar pueden afectar al crecimiento académico del estudiante (Martínez-Arias et al., 2009). Un estudiante que asiste a un centro educativo, independientemente del que sea, el valor añadido permite estudiar el rendimiento que se espera de dicha escuela en comparación con el promedio del resto de escuelas con características afines.

De esta manera una de las variables principales para poder calcular el valor añadido es el rendimiento previo del estudiante, ya que la misma permite controlar factores que pueden afectar al progreso académico de los alumnos (Ballou, Sanders & Wright, 2004; Sanders, 2000). En la práctica existen diversos enfoques para estudiar el valor añadido de acuerdo al número de mediciones (Sanders, 2006; Shin, 2007; Tekwe et al., 2004). Cuando se calcula el aumento de los niveles de logro entre dos puntuaciones, se habla de modelos de ganancia, mientras que si existen más de dos puntos temporales, estos modelos son definidos como modelos de crecimiento, caracterizados porque se estima un punto de partida denominado como intercepto o estatus y una pendiente de crecimiento.

Dichos modelos van desde las más simples estimaciones, donde el crecimiento se calcula como la resta entre la puntuación actual y las anteriores, hasta modelos muy complejos que permiten estudiar el rendimiento de los estudiantes teniendo en cuenta que dicho nivel de logro se ve afectado por la intervención de distintos profesores y, por tanto, es posible aislar el efecto de los distintos docentes sobre el rendimiento estudiantil (modelos de efectos cruzados) (Sanders, 2006).

Todo esto ha originado que los modelos de valor añadido sean considerados como una serie de técnicas complejas que utilizan datos procedentes de evaluaciones educativas realizadas a lo largo del tiempo, para estimar los efectos que la escuela tiene sobre el rendimiento de cada uno de sus estudiantes (McCaffrey et al., 2003a).

Los investigadores inciden en que estos modelos no deben ser la única herramienta para tomar decisiones de alto impacto, como el establecimiento de sanciones o incentivos (Ballou, 2002; McCaffrey, et al., 2003a; Raudenbush, 2004), sino que deben utilizarse con el objetivo de informar, para la introducción de mejoras académicas, ya que dichos modelos cuando se usan con fines sancionadores, podrían afectar a los profesionales educativos en su desempeño profesional y en la participación de estas evaluaciones, lo que conllevaría a la pérdida de información relevante (Yeagley, 2007).

Es necesario profundizar en el estudio de los modelos de valor añadido antes de responsabilizar a nadie de los resultados académicos obtenidos. Aun así, a pesar de no ser perfectos son mejores que los sistemas de evaluación en los cuales, se mide a los estudiantes en un único momento temporal (McCaffrey, et al., 2003a), ya que proporcionan un panorama más completo del rendimiento escolar que las tradicionales evaluaciones de estatus. Sin embargo, no hay acuerdo sobre la precisión de los mismos debido a que dichos modelos se pueden ver afectados por diversos artefactos estadísticos como pueden ser el efecto suelo, efecto techo, el efecto Mateo, el efecto de regresión a la media, etc.

Con el fin de tener una mayor comprensión sobre los principales esquemas analíticos para la medida del crecimiento de los aprendizajes, dicho capítulo I comienza con una pequeña conceptualización sobre las variables de resultados que se miden en las evaluaciones educativas y el impacto que dichos resultados educativos han tenido sobre la rendición de cuentas ("accountability"). También se estudia la importancia del aprendizaje en términos de crecimiento y la diferencia existente entre los modelos de cambio de cohorte, modelos de ganancia y modelos de crecimiento. Tras el estudio de dichos modelos se finaliza este capítulo mediante un análisis detallado de los modelos de valor añadido en educación.

I.1. ¿Qué miden las grandes evaluaciones educativas?, variables de resultado: rendimiento académico y/o competencias

Tal y como se ha comentado en la introducción de este mismo capítulo, las evaluaciones educativas, se centran principalmente en el análisis de los resultados de los estudiantes en términos de rendimiento y/o competencias.

A pesar, de la larga tradición con la que cuenta la evaluación educativa no existe un consenso común sobre la definición de estos dos conceptos, por ello, a continuación, se intenta esclarecer los mismos.

I.1.1. El rendimiento académico

La importancia que los sistemas educativos dan a la consecución de los objetivos propuestos por los grandes organismos políticos, con la finalidad de conseguir financiación, no ser sancionados y satisfacer las demandas sociales, influye en la visión y en la valoración que los centros hacen sobre el rendimiento académico de los estudiantes. Como consecuencia, dicho concepto está muy influenciado no solo por el esfuerzo del sujeto que aprende (Kaczynska, 1965), sino por otros factores que guardan una estrecha relación con las demandas políticas educativas.

El concepto de rendimiento tiene su origen en un modelo económico industrial, cuyo objetivo era el incremento de la calidad y productividad de los trabajadores. Posteriormente dicho concepto se extiende a otros ámbitos, entre ellos, al educativo. Aunque inicialmente, en este campo se identificaba con el término de aprendizaje, ya que ambos conceptos hacen referencia a la práctica docente donde el estudiante aprende y demuestra lo aprendido en una prueba de rendimiento, poco a poco dicho término se ha ido delimitando, pero sin generar una única definición.

Son diversas las ocasiones cuando al hablar de rendimiento académico se alude al nivel de conocimientos del alumno en un determinado dominio, es decir, el rendimiento hace referencia al conocimiento del estudiante acorde a unos objetivos preestablecidos, y en comparación con los resultados obtenidos por su grupo de iguales. En este sentido de productividad, Bloom (1972) define el rendimiento como el producto del trabajo escolar, en la medida en que el estudiante es capaz de aplicar

los conocimientos aprendidos a otras situaciones. Para Martínez-Otero (1996), el rendimiento también es un producto, que se manifiesta en los logros del alumno dentro del centro escolar reflejado generalmente en una calificación.

Otros autores se basan en el concepto de eficacia escolar para definir el rendimiento, en este sentido Pacheco (1970) lo considera como el aprovechamiento escolar que el trabajo educativo produce.

En determinadas ocasiones estas concepciones de productividad y eficacia escolar se han identificado como términos análogos. Sin embargo, la productividad escolar tiene un carácter más economicista y, hace referencia a la eficiencia en la optimización de costes, mientras que, la eficacia escolar se relaciona con los procesos necesarios para la consecución de los objetivos (Murrillo, 2003). Desde esta perspectiva, y en este trabajo, se entiende que una escuela es eficaz "si consigue un desarrollo integral de todos y cada uno de sus alumnos, mejor de lo que sería esperable teniendo en cuenta su rendimiento previo y la situación social, económica y cultural de las familias" (Murillo, 2003, p. 14). En esta misma línea, Castro (2014) considera que la eficacia escolar depende del trabajo conjunto de las familias, el entorno social y las características individuales de los estudiantes. Encontrando que las escuelas más eficaces son aquellas que se centran en labores fundamentalmente pedagógicas como son: la enseñanza de los estudiantes, los logros del aprendizaje, la monitorización del trabajo del alumno y la ayuda al estudiante en la tarea de aprender.

Zapata, De los Reyes, Lewis y Barceló (2009) entienden el rendimiento como un sistema que mide logros y construcción de conocimientos en los estudiantes, desarrollados por la intervención de estrategias y didácticas educativas que son evaluadas a través de métodos cualitativos y cuantitativos en una materia. Esta concepción de rendimiento más vinculada a la medición y evaluación fue reflejada en Carabaña (1987), quien otorgó una gran importancia a la validez de las distintas medidas del rendimiento, puesto que consideraba que el éxito de definir bien el rendimiento dependía del resultado de sus mediciones sociales y académicamente relevantes.

Por todo lo comentado, y siguiendo la idea de Zapata et al. (2014), parece que el concepto de rendimiento académico esta inacabado y se va construyendo a partir de distintas definiciones que van integrando diferentes elementos que conforman su carácter multidimensional. Lo que pone de manifiesto la complejidad del término rendimiento, no sólo a nivel conceptual sino también a nivel metodológico.

En este trabajo se considera que el rendimiento académico, supone una valoración acerca de la adquisición de conocimientos, destrezas y actitudes por parte del estudiante a lo largo de su etapa escolar, siendo el reflejo de la actividad educativa.

I.1.2. Competencias educativas

El rendimiento no es la única forma de medir el desempeño académico de los estudiantes, en otras ocasiones, los resultados de las evaluaciones se miden en términos de competencias.

Las sociedades modernas se han transformado drásticamente debido al continuo cambio global que han experimentado. Los estudiantes necesitan ser flexibles a las demandas sociales para que puedan adaptarse a las sociedades complejas y multiculturales de hoy en día. Para apoyar las necesidades de estos nuevos estudiantes, la educación está cambiando su enfoque tradicional basado en la transmisión de conocimientos aislados, a la adquisición de competencias complejas demandadas socialmente. Los alumnos tienen que desarrollar una serie de habilidades para el aprendizaje y la obtención de información, a partir de una serie de fuentes disponibles en las sociedades actuales.

El término competencia, no apareció directamente en el ámbito educativo, sino que tiene su origen en el área empresarial (Denyer, Furnémount, Poulain & Vanloubbeeck, 2007) de ahí, que dicho concepto sea un tanto controvertido en el campo educativo, puesto que se consideraba un tanto "mercantilista" que podía desvirtuar los planteamientos pedagógicos propios de los procesos de enseñanza-aprendizaje.

Se considera a McClelland el divulgador del concepto de competencia, ya que es en su artículo *Testing for Competence Rather Than for Intelligence* (1973), donde por primera vez se define las competencias como una habilidad de trabajo que,

supone un desempeño competente en dicho campo (McClelland, 1998). El enfoque de competencias de McClelland (1973) abogó por el uso de este concepto, en lugar del término de inteligencia en las pruebas, y mostró cómo identificar las competencias a través de entrevistas de eventos conductuales. En este sentido, las competencias se adquieren a través de la capacitación y el desarrollo, basándose en la descripción del comportamiento o un desempeño observable in situ.

Así, este término procedente del campo laboral, ha afectado a la propia definición del concepto de competencia dentro del campo educativo, ya que una razón importante de la popularidad del mismo es la expectativa de reducir la brecha entre el mercado de trabajo y la educación (Denyer et al., 2007), a través de un sistema educativo basado en competencias. La idea subyacente es que la formación profesional debe ayudar a los estudiantes a adquirir las competencias necesarias en sus futuras profesiones y en la sociedad en su conjunto. Además, al trabajar como profesionales, deben continuar desarrollando sus competencias para que puedan reaccionar y anticipar desarrollos futuros en su trabajo y fuera del mismo (Jenewein, Knauth & Zülch, 2002).

Por lo tanto, se reconoce cada vez más la necesidad de que la formación se oriente al desarrollo de competencias y no sólo a la adquisición de conocimientos estáticos. Se debe enfatizar las capacidades y no las calificaciones, ya que las capacidades mejoran el vínculo entre la educación y el mercado de trabajo (Mulder, 2001). Además, según este mismo autor, el reconocimiento de las competencias adquiridas de manera informal y las adquiridas fuera del sistema educativo, también desempeñan un papel importante en la formación integral del estudiante.

En este contexto, la educación y la formación basada en las competencias tiene una larga historia. Achtenhagen y Grubb (2001) rastrean los primeros enfoques de análisis de tareas de competencias en Moscú en la década de 1860. Nijhof (2003) también enfatiza una larga historia de la educación basada en la competencia, refiriéndose al enfoque de Bobbitt en la década de 1920, mediante el cual se realiza un análisis científico de las acciones humanas para identificar las habilidades subyacentes necesarias para el alto rendimiento basado en las competencias.

Todo esto ha ido originando que, con el paso del tiempo, la educación basada en las competencias esté adquiriendo cada vez más importancia dentro del marco educativo, con el fin de abordar las deficiencias entre la educación y la formación profesional.

La importancia de definir el concepto de competencia se desprende del hecho de que los planes de estudios y las evaluaciones están, en gran medida, determinados por los resultados de aprendizaje que se quiere que los estudiantes logren y que, a su vez, se encuentran influenciados por las concepciones de la competencia (Lizzio y Wilson, 2004).

Eraut (1994), ofrece una visión general del concepto de competencia donde distingue tres tradiciones principales que se han mantenido hasta hoy en dia. La primera de ellas dentro de la psicología conductista, centrada en especificaciones muy detalladas del comportamiento competente, basandose exclusivamente en el proceso técnico de análisis de tareas, pero dejando de lado las dimensiones sociales y políticas del desarrollo de la competencia. En segundo lugar, los enfoques genéricos de la competencia, que tuvieron como objetivo principal identificar las cualidades generales vinculadas con el rendimiento excelente en el trabajo y, se centraron más en la selección que en el entrenamiento o la enseñanza. Y, en tercer lugar, la definición de competencia basada en construcciones derivadas de la psicología cognitiva. Respecto a esta tradición, afirma que los investigadores trataron de distinguir entre competencia y desempeño, para ello se basaban fundamentalmente en aspectos relacionados con las implicaciones metodológicas que cada uno de estos constructos tenían dentro de las evaluaciones (Chomsky, 2006 & Messick ,1984).

Aunque, como bien señala De la Orden (2011) y Merriënboer, Van der Klink y Hendriks (2002) la competencia es un término polisémico, de carácter multidimensional que abarca una gran variedad de aspectos, como consecuencia de que existen muchas definiciones y poco consenso sobre el significado de estos conceptos entre los muchos investigadores y autores (Actenhagen, 2005; Chomsky, 2006; Eraut, 1994; Jong, 2003; Lizzio & Wilson, 2004; Mulder, 2001; Winterton Delamare Le-Deist & Stringfellow, 2005), quizás una de las definiciones más precisas es la proporcionada por MECD (2015) donde se entienden las competencias

educativas como un conjunto de destrezas, conocimientos, habilidades, actitudes, emociones y valores éticos, contextualizados en la realidad del alumnado, con el fin de promover su desarrollo y realización personal, así como integración social, en el ejercicio de una ciudadanía activa e integradora y con capacidad para aprender permanentemente. Partiendo de esta idea, en dicho trabajo se entiende que la competencia educativa es la capacidad del estudiante para generalizar los conocimientos, habilidades, actitudes y destrezas aprendidas en un contexto determinado a situaciones afines, pero no iguales, que permitan al individuo un desarrollo tanto personal como profesional.

I.2. Rendición de cuentas (accountability)

En las últimas décadas ha habido un interés incipiente, especialmente por parte de las políticas educativas, en usar los resultados obtenidos en términos de rendimiento y/o competencia, en las distintas evaluaciones educativas realizadas por los estudiantes, como un medio de rendición de cuentas (*accountability*) por parte de las escuelas. La creciente demanda de responsabilizar a las escuelas por los resultados de los estudiantes ha adquirido cada vez una mayor atención. En general, se reconoce que un sistema escolar y una escuela eficaz proporcionan un ambiente estimulante para el aprendizaje académico de los estudiantes, pero no se puede esperar que todos los estudiantes de distintos y mismos centros rindan de igual manera.

El fenómeno *accountability* empieza adquirir importancia en Estados Unidos con la Ley *No Child Left Behind* (NCLB) que se promulgó en el año 2001. Dicha ley exigía que todos los estudiantes de las escuelas públicas desde 3º a 8º curso fueran competentes en Matemáticas y Comprensión Lectora. Para ello, se difundía un sistema de recompensas dirigido a aquellas escuelas que conseguían que sus estudiantes tuvieran un progreso anual adecuado (*Adequate Yearly Progress,* AYP) y sancionaba a aquellas donde los resultados de las pruebas de los estudiantes no eran los esperados. Las sanciones eran tales que en el caso de que una escuela con un gran número de estudiantes procedentes de familias con bajos ingresos no cumpliera con el AYP, la ley autorizaba al estado a cambiar al equipo que dirigía el centro o incluso a cerrar la escuela. Dichas penalizaciones dependían de cada estado, por ejemplo, en Texas si las escuelas no conseguían los objetivos propuestos se les

retiraba parte de la financiación. Estas consecuencias llevaban a que muchas escuelas se centraran en preparar a sus estudiantes para pasar este tipo de pruebas o incluso a sacar a los alumnos menos preparados de las aulas para que no realizaran las mismas.

En esta ley, el progreso adecuado (AYP) se operacionalizaba como una comparación entre la tasa de crecimiento de una escuela en un momento determinado, con el mínimo de cambio exigido para dicha escuela en ese curso concreto. Dicho análisis proporcionaba una doble información, por un lado, notificaba si la escuela era eficaz y, por otro lado, si a nivel individual el estudiante era competente. Un AYP positivo indicaba que los estudiantes habían obtenido un nivel adecuado de aprovechamiento en las evaluaciones, según los estándares académicos prefijados por el estado, mientras que un AYP negativo indicaba un nivel de aprovechamiento insuficiente y, por lo tanto, suponía una serie de penalizaciones.

Las escuelas debían publicar los resultados obtenidos e informar a las familias de los mismos. Ya que la idea que subyacía detrás de este fenómeno de rendición de cuentas es que, tanto los padres, como los ciudadanos que pagaban los impuestos, tenían derecho a saber que se estaba haciendo con su inversión.

Asimismo, los centros educativos debían de informar de los resultados obtenidos por los estudiantes que estaban recibiendo educación especial. Estos alumnos, a pesar de sus posibles dificultades, también debían lograr el nivel de competencia o aptitud mínimo establecido en las pruebas, ya que se consideraba que la única manera de que las escuelas mejoraran y fueran eficaces y eficientes era que todas, independientemente del tipo de alumnado y centro, alcanzaran una serie de objetivos.

La primera aplicación de esta evaluación estatal, originada con la ley NCBL, se centró en el estudio del estado actual del rendimiento de los estudiantes (evaluación de estatus). El enfoque del estado actual compara diferentes cohortes de estudiantes en un solo momento temporal (Doran & Izumi, 2004). Para ello, simplemente utilizaban el porcentaje de estudiantes que aprobaban la prueba estatal al final del año escolar. A pesar de que los educadores reconocían que una puntuación de logro en un solo momento temporal, no siempre era una manera útil

de estimar los efectos de la escuela sobre el desempeño del estudiante, era un modelo ampliamente utilizado.

Aunque estos métodos utilizados en la ley NCBL se localizan en el corazón del sistema estatal de rendición de cuentas, hay distintas razones por las cuales eran y son inválidos e inapropiados a la hora de realizar comparaciones escolares.

En primer lugar, estos métodos de estado actual no tenían en cuenta que las diferencias entre las escuelas podían deberse a variables estudiantiles y escolares que influyen en los resultados de las pruebas, como pueden ser factores socioeconómicos o rendimiento previo. No había asignación al azar de los estudiantes a las escuelas. Todo esto daba lugar a comparaciones injustas entre las escuelas desfavorecidas y favorecidas en términos de estatus socioeconómico (Doran & Izumi, 2004). En segundo lugar, los métodos de estado actual son acumulativos; reflejan el impacto del aprendizaje obtenido de todas las escuelas anteriores en las puntuaciones de rendimiento de los estudiantes (Doran & Izumi, 2004). No se puede considerar a la última escuela que ha asistido el alumno como la responsable de su buena o mala puntuación en la prueba, si un estudiante ha pasado por distintos centros durante su educación. Además, como apunta Ballou et al. (2004), el hecho de que se considere a las escuelas responsables de los niveles medios de logro, no tiene sentido cuando los estudiantes entran en las escuelas con grandes diferencias en el rendimiento.

Otro argumento en contra del estudio del estado actual del rendimiento dentro de la rendición de cuentas, es que las escuelas con una gran proporción de estudiantes académicamente desfavorecidos son injustamente juzgados, ya que, como bien se ha comentado en este mismo capítulo, se les exige los mismos estándares que las escuelas más favorecidas, sin tener en cuenta factores no escolares (por ejemplo, estatus socioeconómico) (Braun, 2005; Kane & Staiger, 2002; Raudenbush, 2004).

Bajo este contexto, habría que reflexionar si los métodos utilizados ya en la ley NCLB para medir el progreso escolar (evaluaciones de estatus), captan de manera fiable el impacto que tienen las escuelas en los estudiantes y si están sesgados en contra de las escuelas que atienden a estudiantes procedentes de entornos desfavorecidos.

Un tema de gran importancia en la medición del rendimiento escolar, es el exceso de confianza que los sistemas de rendición de cuentas y, en concreto, la ley NCBL, dan a la comparación de cohortes sucesivas de estudiantes como una medida de "cambio". Por ejemplo, comparan el rendimiento promedio de alumnos de 6º grado en el año 2001, con el nivel de logro de los alumnos de ese mismo grado en el año siguiente, con la finalidad de medir y evaluar la escuela centrándose en el cambio producido de año a año en las puntuaciones de rendimiento de los estudiantes. La comparación de sucesivas cohortes de estudiantes, permite a los estados evaluar a las escuelas en término de ganancia, en lugar de niveles de rendimiento absolutos. Sin embargo, el uso de diferentes cohortes de estudiantes para medir el progreso de la escuela o la mejora de las mismas, es problemático para los propósitos de la evaluación, ya que diversas investigaciones demuestran que las estimaciones de ganancia de año a año en distintas cohortes, están afectadas en gran medida por la variación del muestreo, el error de medición y otros factores que no están asociados con el tamaño de la escuela o la práctica escolar (Kane & Staiger, 2002; Linn & Haug, 2002).

Esto, junto con muchas otras críticas realizadas a los métodos de rendición de cuentas promulgados con la ley NCBL, ha hecho que los sistemas de rendición de cuentas basados en el estado actual sean considerados como métodos defectuosos, que dan lugar a juicios inexactos sobre la calidad de la escuela.

La ley NCBL fue abolida en el año 2015 cuando entró en vigor una nueva ley, *Ley Cada Estudiante Triunfa* (*Every Student Succeeds Act* o ESSA) que mantuvo ciertas partes de NCLB y revocó otras. Entre los nuevos cambios introducidos respecto a la ley anterior, destaca que cada estado establece los objetivos y evalúa los logros de las escuelas. Del mismo modo, los estados deben desarrollar planes de mejora para aquellas escuelas que tienen problemas o bien para un grupo determinado de estudiantes cuyo rendimiento no es el adecuado. Además, a la hora de evaluar a las escuelas no solo se consideran las puntuaciones de las mismas, sino que se tiene en cuenta otros aspectos como la seguridad escolar o el acceso a cursos avanzados, aunque la puntuación de desempeño sigue siendo el indicador más importante y, se siguen basando en evaluaciones de estatus. En dicha ley se mantiene un sistema de penalizaciones por no alcanzar los estándares establecidos,

pero el gobierno ya no exige que todos los estudiantes aprueben los exámenes estatales y, no es necesario que los estados logren las metas preestablecidas, lo que elimina las sanciones tan severas que eran impuestas a nivel federal.

Lo más importante de la ley ESSA es que permite que cada distrito desarrolle su propio programa educativo acorde a la realidad específica de su población, hecho que deja espacio a la diversidad y multiculturalidad; no como la ley anterior (NCBL), donde era necesario cumplir unos estándares nacionales independientemente de las características de los estudiantes y de la escuela en general.

Esta legislación ha sufrido modificaciones, ya que con el actual gobierno de Donald Trump se derogan las normas que exigían a los centros y programas de entrenamiento de profesores, a rendir cuentas al gobierno federal de los resultados obtenidos. La finalidad de Trump con la introducción de estos cambios, según apunta él mismo, es garantizar que los estados y distritos se centren en la mejora de los resultados y en la medición del progreso de los estudiantes, proporcionando una mayor libertad a las escuelas (Jiménez, 2017).

A medida que las deficiencias de las evaluaciones de estatus se han hecho evidentes, se han planteado una forma alternativa de evaluar la efectividad de la escuela, que ha ganado terreno en el sistema de rendición de cuentas. Este nuevo método se centra en estudiar el cambio o crecimiento del alumno en distintos momentos temporales, lo que permite analizar cómo los grupos de una misma cohorte han aumentado su conocimiento, midiendo el progreso del estudiante a lo largo del tiempo y estudiando los cambios producidos de una medición a la siguiente; algo mucho más razonable, ya que el aprendizaje debe ser entendido como un fenómeno dinámico y no estático. Esta curva de crecimiento desarrollada a lo largo del tiempo, se estudia, sobre todo, mediante los denominados modelos de crecimiento, diseñados fundamentalmente para generar estimaciones que permiten analizar el progreso del estudiante o del centro, en diversas ocasiones de medida (Doran & Izumi, 2004).

Por lo comentado hasta el momento parece que la rendición de cuentas se ha basado exclusivamente en sistemas de alto impacto ("*high-stakes*"), donde los resultados obtenidos en las evaluaciones se utilizaban con fines sancionadores, de tal manera que, como en el caso de la ley NCBL (2001), si las escuelas no conseguían

que sus estudiantes alcanzaran los niveles de logro establecidos eran penalizadas. Pero la rendición de cuentas también puede centrarse en un sistema de bajo impacto *("low-stakes"*), cuyo objetivo es realizar un diagnóstico de la situación académica, para analizar cómo están funcionando las escuelas, cual es el rendimiento del alumnado, etc. Todo ello con fines de informar y mejorar la situación escolar en la que se encuentran los centros, sin intenciones de penalizar a los mismos.

Las evaluaciones para rendir cuentas también permiten comparaciones entre escuelas o sistemas educativos, a través del establecimiento de rankings que, en cierta medida, es otra forma de sancionar a los centros educativos de manera indirecta. Se distinguen fundamentalmente dos tipos de *rankings*: *los de resultados brutos, q*ue hacen referencia a la puntuación media extraída de los resultados de logro; por lo que dependen de factores socioeconómicos y tienen en cuenta la interacción entre los estudiantes y, los *rankings de valor añadido contextual* (ver Ray, Helen & Tanya, 2009), donde en las puntuaciones de logro se aísla el efecto del centro del resto de factores, por lo que permite estudiar lo que aporta realmente el centro a la tasa de rendimiento del estudiante (Calero & Choi, 2012). En este tipo de ranking el posicionamiento de las escuelas, se realiza mediante el estudio del residuo del centro obtenido en los modelos de valor añadido

El uso de rankings como medio de comparación, requiere de una metodología muy minuciosa que, si no se tiene en cuenta puede conducir a una serie de errores en el establecimiento de los mismos. Aunque generalmente las puntuaciones utilizadas en los distintos sistemas educativos se encuentran en una escala común, existen una serie de componentes diferenciales que pueden afectar a los resultados y, que a menudo pasan desapercibidos a la hora de realizar estas clasificaciones, como son, por ejemplo, la diversidad estudiantil, variables sociodemográficas del centro y del alumnado, el tamaño de las aulas, etc. Pero además en el estudio de los ránkings, existe una tendencia en intentar dar respuesta al por qué unos centros o sistemas educativos se encuentran mejor situados que otros, dejando de lado la importancia de si dichas comparaciones reflejan realmente la situación real de un centro.

Para garantizar la idoneidad de los resultados, la evaluación debe proporcionar comparaciones fiables de los mismos, mediante el empleo de una metodología adecuada que abarca, desde la elaboración de un instrumento de medida válido y fiable, hasta la aplicación de un proceso de análisis de datos específico.

Pero incluso utilizando una metodología tan exhaustiva como pueden ser los modelos de valor añadido, para el establecimiento de dichas clasificaciones, no se está exento de que los resultados se vean afectados por ciertos artefactos estadísticos, difíciles de identificar y paliar, como es el conocido, pero poco entendido, efecto de regresión a la media. Como se estudiará más adelante, dicho artefacto se encuentra siempre presente entre la primera medición y el resto de ocasiones de medida, a pesar de que rara vez se tiene en cuenta.

I.3. La importancia del crecimiento en el aprendizaje

La noción de aprendizaje implica cambio y crecimiento. La medida del cambio ofrece información útil para evaluar el aprendizaje. Solo midiendo el cambio individual es posible informar sobre el crecimiento de cada persona y, por lo tanto, evaluar la efectividad de los sistemas educativos (Willett, 1994).

Las medidas del rendimiento medio, como la puntuación media de los logros de los estudiantes o la tasa media de crecimiento, tienden a ser los indicadores más comunes para el seguimiento del cambio escolar. Pero los promedios tienden a simplificar en exceso lo que está pasando dentro de una escuela, por ello, mucha más información sobre el desempeño escolar se puede obtener mediante estimaciones de la tasa de crecimiento. En este contexto los modelos de valor añadido son a menudo los métodos estadísticos más utilizados para medir el crecimiento del estudiante con fines de rendición de cuentas.

La medición del cambio es un tema familiar en las ciencias del comportamiento; documentos sobre este tema han estado apareciendo en la literatura durante más de 50 años.

Ya en 1958, Lord afirmó que el propósito de los datos sobre el cambio individual es hacer inferencias acerca de la verdadera ganancia del examinado, que se define como la diferencia entre sus puntuaciones verdaderas en el pre-test y en

el post-test. Unos años más tarde, Harris (1967) escribió un libro titulado "*Los problemas en la medida del cambio*", donde plantea una serie de preguntas importantes sobre la medición y el análisis del cambio, además, sugirió algunas nuevas técnicas como el análisis de series temporales. Desde 1967, ha habido muchos desarrollos analíticos y conceptuales de lo que es el cambio, entre ellos se puede destacar una serie de volúmenes publicados donde se reflejan los aspectos más relevantes del estudio del cambio en educación (por ejemplo, Betebenner, 2009; Betebenner & Linn, 2010; Collins & Horn, 1991; Collins & Sayer, 2001; Curran, Lee, Howard, Lane & MacCallum, 2012; Muniz et al., 2017; Rovine & VonEye, 1991; VonEye, 1990).

El cambio en su forma más básica se analiza como una simple transformación de las puntuaciones en los test de un momento a otro. Modelos más complejos incorporan técnicas estadísticas que utilizan información longitudinal, no una simple relación pretest-postest, y pueden tener en cuenta diferentes factores del contexto del estudiante e incluso los efectos previos producidos por otras escuelas.

El cambio, también se puede estudiar de otras maneras, por ejemplo, como las modificaciones producidas a lo largo de una escala de puntuaciones obtenidas mediante test de rendimiento, como la proporción de estudiantes que alcanzan o superan un determinado estándar, como la diferencia entre el crecimiento observado y el crecimiento esperado, etc. Por lo tanto, diferentes tipos de ganancia o crecimiento dan lugar a diferentes formas de medir el cambio en educación, aunque la idea de modelar el crecimiento subyacente de un individuo, consiste básicamente en estimar la línea de base de la persona (punto de partida) y la trayectoria (pendiente) en mediciones consecutivas tomadas en un mismo individuo.

En este sentido, parece claro que el estudio del cambio en la investigación educativa permite analizar el crecimiento de los estudiantes cuando adquieren nuevas habilidades, cuando aprenden algo nuevo, cuando crecen intelectualmente y físicamente y, sus actitudes e intereses se desarrollan produciendo un cambio en los mismos. Así, hay que conocer y medir este cambio para tener una idea sobre el desarrollo y el crecimiento tanto de los individuos como de las escuelas.

Para medir el cambio, los modelos de crecimiento o valor añadido son vistos cada vez más como un enfoque estadísticamente válido, que permite cuantificar el cambio de los aprendizajes a lo largo del tiempo, ya que, en los últimos años, los métodos de modelado de crecimiento individual, han aprovechado la riqueza multivariante continua de los datos, para proporcionar mejores métodos que respondan a las preguntas sobre diferencias interindividuales sistemáticas en el cambio (Bryk, 1977; Bryk & Raudenbush, 1987; Rogosa & Willett, 1985; Willett, 1988, 1994).

I.4. Modelos de cambio de cohorte, modelos de ganancia y modelos de crecimiento

Previo a la descripción de los modelos de valor añadido, es necesario explicar los modelos de ganancia y de crecimiento en los que se basa para la estimación del residuo de crecimiento, así como los modelos de cambio de cohorte por la importancia que han tenido en la evaluación educativa.

El logro de los aprendizajes se puede estudiar de dos maneras distintas, o bien como estatus o bien como cambio o crecimiento. Mientras que, para el estudio del estatus se suelen emplear los modelos de cambio de cohorte, a través de los cuales, se pretende estudiar el rendimiento conseguido en un determinado curso midiendo a distintos alumnos en diferentes momentos, generalmente años consecutivos, en el estudio del cambio se suelen utilizar los modelos de ganancia o de crecimiento.

Se alude al término ganancia cuando se estima el cambio producido en el rendimiento en dos momentos temporales (pre-test, post-test). Sin embargo, para estimar el crecimiento se requiere de más de dos puntuaciones de logro.

Aunque los modelos de ganancia cuentan con estimaciones fiables, son los modelos de crecimiento los que proporcionan más información sobre el cambio que se produce en los estudiantes y, el efecto que determinados predictores, a nivel de estudiante y/o de centro tienen sobre el crecimiento (Rogosa & Willet 1983). Dichos predictores en los modelos de ganancia pierden fuerza (Willet, 1989; Willet, 1994) por ello, según determinados autores (por ejemplo, Steven & Zvoch, 2006; Thum, 2006; Willet, 1989; Willet, 1994), para evaluar el progreso a nivel individual de los

estudiantes y de escuelas, son más adecuados los modelos de crecimiento, aunque suponen un mayor coste.

En todos los modelos que se desarrollan a continuación se pueden considerar los efectos de los centros como efectos fijos o aleatorios, en este trabajo para no ser repetitivos se realiza una aproximación aleatoria, ya que los modelos de efectos fijos son más comunes en la regresión simple o múltiple. La principal diferencia con el modelo fijo es que la varianza entre centros desaparece.

I.4.1. Modelos de cambio cohorte a cohorte

Estos modelos se consideran modelos de estatus, ya que, a diferencia de los modelos de ganancia o crecimiento, no se mide a un mismo sujeto en distintas ocasiones temporales, sino que generalmente se realizan dos mediciones del logro en un mismo curso a diferentes estudiantes, es decir, se estudia el logro conseguido en un grado concreto, midiendo a distintos alumnos de ese curso en distintos años.

Estas mediciones a nivel escolar tan sólo permiten conocer si de una medición a otra ha habido modificaciones en el número de estudiantes que se encontraban en un determinado nivel académico. De esta manera, dichos modelos permiten comparar los resultados obtenidos por los centros educativos en un curso determinado, con los que obtienen los alumnos de ese mismo curso al año siguiente, pero no proporcionan información de la trayectoria de los estudiantes a lo largo del tiempo, ya que los estudiantes no son los mismos.

En estos modelos se suelen realizar mediciones con carácter anual de un grado determinado y se analizan las puntuaciones de logro medio, respecto a unos estándares establecidos previamente. Así, el progreso se refiere al número de alumnos que consiguen el nivel de logro preestablecido en un grado especifico.

Este tipo de modelos es el que se estableció en Estados Unidos, bajo la ley NCBL (estudiada con anterioridad), con el apelativo de Progreso Anual Adecuado (AYP, por sus siglas en inglés), donde se pretendía que todas las escuelas consiguieran las metas establecidas a nivel federal y estatal sobre aprovechamiento académico. El problema de dichos modelos, tal y como se ha descrito, es que no refleja la situación real del centro, ya que como apuntan Choi, Goldschmidt y Yamashiro(2006), los resultados no tienen en cuenta el rendimiento previo y otras

variables, como el nivel socioeconómico, sin las cuales es dificil conocer cómo afectan los procesos de enseñanza-prendizaje llevados acabo en la escuela sobre el nivel competencial del alumno. Además, al no analizar el porcentaje de estudiantes que no alcanzan los objetivos preestablecidos, se deja de lado a los estudiantes que no consiguen alcanzar un nivel determinado, de tal manera que se pierde informacion relevante sobre la situacion académica del centro.

En definitiva, estos modelos que solo estudian los cambios en el logro que se producen de una cohorte a otra, proporcionan una información muy limitada del cambio y crecimiento que se origina en la escuela. Solo permiten estudiar la eficacia que se produce en un determinado curso de un año a otro, sin tener en cuenta una serie de variables que pueden afectar a los resultados obtenidos.

Sin embargo, las evaluaciones que miden a los mismos estudiantes a lo largo del tiempo proporcionan mucha mas información, ya que no se evalúa la situación académica en un único momento temporal sino, que permite modelar la trayectoria del cambio y del crecimiento a lo largo del tiempo, reflejando con mayor exactitud cómo se produce realmente el aprendizaje de los estudiantes en las escuelas.

I.4.2. Modelos de ganancia

Los modelos de ganancia se caracterízan porque analizan a una misma cohorte en dos momentos distintos, generalmente en dos cursos sucesivos, con la finalidad de analizar si los estudiantes han progresado en términos promedios.

Estos modelos asumen que una sola medida de logro no es suficiente para analizar el progreso del estudiante y, por ello, se estudia el cambio entre dos puntos temporales, un pre-test y un post-test, donde se analiza el incremento que se produce entre el estatus inicial del estudiante o grupos de estudiantes, antes de un periodo de aprendizaje y su rendimiento después de ese periodo. De esta manera, el cambio no es considerado como un desarrollo a lo largo del tiempo, sino como la cantidad de mejora (ganancia) en la variable de logro que se produce entre el pre-test y el post-test.

En los modelos de valor añadido, los análisis de ganancia son considerados como el cambio que se produce en los estudiantes entre el pre-test y el post-test, a consecuencia del efecto que el docente tiene sobre el rendimiento del estudiante.

Así, para cada alumno la puntuación inicial es sacada de la puntuación actual y de esta manera, se calcula una medida del cambio en el rendimiento académico de los estudiantes en una asignatura y/o aula concreta. La media de las ganancias de todos los estudiantes de una misma aula se compara con otros grupos de alumnos, siendo posible la comparación de diferentes docentes dentro de una misma escuela. El valor añadido de la escuela se considera la diferencia entre la ganancia de un centro determinado y la del grupo de comparación, como por ejemplo la media del distrito, la media global, etc. (Navarro, 2013).

Dentro de los modelos de ganancia se pueden distinguir tres tipos de ganancia (Hibpshman, 2004):

A) Ganancia bruta:

La ganancia bruta se refiere al incremento que se produce entre las dos estimaciones realizadas. Generalmente la ganancia bruta es una medida directa de la distancia entre el nivel de logro actual ($t = 1, Y_1$) de un sujeto o grupos de sujetos y el nivel inicial ($t = 0, Y_0$), es decir, ($Y_1 - Y_0$) (Martinez-Arias et al., 2009). Esta puntuación es utilizada en los análisis de valor añadido como una variable dependiente en modelos de efectos fijos o aleatorios. Según Navarro (2013), si en dichos modelos se añaden variables contextuales, como predictores de los alumnos y los centros y se realizan estimaciones a nivel de centro, se pueden considerar un modelo de valor añadido.

Siguiendo a este mismo autor, un modelo de ganancia multinivel (nivel de estudiantes y escuelas), donde la puntuación de ganancia es la variable criterio y los efectos de los centros son variables aleatorias, se puede formular de la siguiente manera:

Ec. I.1

Nivel 1 (estudiantes): $G_{ij} = \beta_{0j} + \beta_{qj}X_{qij} + e_{ij}$

Nivel 2 (escuelas): $\beta_{0j} = \beta_{00} + \beta_q W_{sj} + r_{0j}$

$$\beta_{qj} = \beta_{q0} + r_{qj}$$

Donde el nivel de escuelas modela los interceptos de los centros obtenidos en el nivel de estudiante. Si se sustituyen las ecuaciones de los niveles anteriores, el modelo queda definido de la siguiente manera:

$$G_{ij} = \beta_{00} + \beta_q W_{sj} + (\beta_{q0} + r_{qj})X_{qij} + r_{0j} + e_{ij} \qquad \textbf{Ec.I.2}$$

La ecuación anterior de regresión (Ec.I.2) con dos niveles, muestra la ganancia bruta de los distintos alumnos (G_{ij}) que se encuentra ajustada con sus posibles predictores de contexto (X_{qij}) (por ejemplo, sexo, nivel socioeconómico, etc.). De esta manera, β_{0j} es la ganancia media de todos los alumnos del centro *j* y β_{qj} es la variación sobre la ganancia del centro *j* causada por las variables contextuales de los sujetos.

Siguiendo a Navarro (2013), cada uno de los coeficientes del nivel de estudiante pasa a ser una ecuación del nivel de centro. De esta manera β_{00} hace referencia a la ganancia media global para todos los centros; β_q es la variación producida como consecuencia de la introducción de cada una de las variables predictoras del contexto escolar (W_{sj}) en la ganancia media, y β_{q0} es la variación que se produce sobre la ganancia global de cada predictor de primer nivel. Además, los residuos de ambos niveles se consideran que son independientes y se distribuyen de manera normal (ver ecuación I.3). El residuo del nivel de centro se caracteriza por tener una matriz de varianzas-covarianzas R que depende de los predictores de la ganancia introducidos en el modelo.

$$\begin{aligned} e_{ij} &\sim N(0, \sigma_e^2) \\ r_{qj} &\sim N(0, R) \end{aligned} \qquad \textbf{Ec. I.3}$$

Finalmente, r_{0j} alude a la desviación de una escuela determinada *j* respecto a la media global. Así, el efecto que produce sobre la ganancia de asistir a un determinado centro *j*, cuando se aíslan determinados factores contextuales, es $\beta_{00} + r_{0j}$, que es el valor añadido de la escuela.

Una limitación de las medidas de ganancia bruta es que fracasan en la consideración de los distintos factores individuales y contextuales, dando lugar a estimaciones "injustas". Esta operacionalización se puede mejorar mediante el uso

de covariables en el post-test, considerando el rendimiento previo como la covariable principal. La distancia se formularia de la siguiente manera, $Y_{i1} - E(Y_{i1}|Y_{i0})$ y se denomina ganancia residual (Martínez- Arias et al., 2009).

B) Ganancia residual

En este tipo de ganancia la puntuación del post-test se utiliza como variable dependiente y como covariable la puntuación del pre-test.

Con el objetivo de evitar la correlación entre la ganancia estimada y el pre-test que se da en la ganancia bruta se opta por una ganancia de tipo residual (Willett, 1994), donde se considera al pre-test como la covariable del post-test. Siguiendo a Navarro (2013), la regresión de Y_1 sobre Y_0 da lugar a $E(Y_1|Y_0)$ y la ganancia se define de la siguiente manera.

$$Y_1 - E(Y_1|Y_0) \qquad \textbf{Ec. I.4}$$

En consecuencia,

$$Y_{i1} = \beta_0 + \beta_1 Y_{i0} + e_i$$
$$e_i \sim N(0, \sigma_e^2) \qquad \textbf{Ec. I.5}$$

En la formulación anterior (ecuación I.5) el término e_i se refiere a la ganancia residual que se obtiene cuando ya se ha introducido como covariable la puntuación de logro, la puntuación previa obtenida por el estudiante (Y_{i0}). La ganancia obtenida se refiere al residuo que se ha producido por el análisis de regresión, es decir, la puntuación observada menos la puntuación estimada, donde el rendimiento previo se utiliza como principal covariable del análisis. Estos modelos no informan de la cantidad en la que un individuo cambia en un atributo determinado, sino de cuál es la puntuación de logro del estudiante eliminando el efecto del rendimiento previo, aun así, una ventaja de estos modelos es que no requieren de puntuaciones escaladas.

La formulación de dicho modelo es similar al modelo de efectos aleatorios comentado para de la ganancia bruta (ver ecuación I.1), pero en este caso se utiliza únicamente el post-test como variable criterio y se añade el pre-test y una serie de

variables contextuales como variables predictores, tal y como se puede observar a continuación:

Ec. I.6

Nivel 1 (estudiantes):

$$Y_{1ij} = \beta_{0j} + \beta_1 Y_{0ij} + \beta_q W_{qij} + e_{ij}$$

Nivel 2 (escuelas):

$$\beta_{0j} = \beta_{00} + r_{0j}$$

$$e_{ij} \sim N(0, \sigma_e^2)$$

$$r_{0j} \sim N(0, \sigma_r^2)$$

En la ecuación anterior (Ec. I.6), β_{0j} se refiere al valor del rendimiento medio de los estudiantes (i) del centro *j* cuando los predictores toman un valor de 0, de esta manera la interpretación del mismo depende de las covariables que se introduzcan en el modelo y del significado que se dé al valor 0 del pretest, por ejemplo si se centra respecto a la media global, β_{0j} representaría la media en el post-test de los alumnos de la escuela *j* con un determinado nivel medio de rendimiento previo; β_1 y β_q se refieren a los coeficientes de regresión que muestran los efectos del pre-test y las covariables de contexto (W). Los residuales de las ecuaciones (e_{ij} y r_{0j}) son considerados independientes y con distribución normal. En la ecuación de los centros, los interceptos de los mismos obtenidos en el primer nivel tienen una distribución aleatoria entorno a la media de todos los centros (β_{00}) así, las desviaciones respecto a dicha media, los residuos del nivel de centro (r_{0j}), son las estimaciones del valor añadido de las escuelas.

C) Ganancia estimada

La ganancia estimada queda expresada de la siguiente manera $E(Y_1) - E(Y_0)$ y supone una mejora respecto a los modelos de ganancia anteriores, ya que no se ven afectados por artefactos de diseño (Castro & Gaviria, 2009). Esta ganancia estimada se utiliza como variables de resultado de las estimaciones del pre-test (rendimiento inicial) y del post-test (rendimiento actual), de tal manera que la ganancia se estima de la siguiente manera:

Ec. I.7

$$Y_{0i} = \beta_{00} + e_{0i}$$

$$Y_{1i} = \beta_{10} + e_{1i}$$

En la ecuación anterior (Ec.I.7), β_{00} y β_{10} se refiere a las medias del estudiante en el rendimiento inicial y en el rendimiento actual, respectivamente, e_i son las desviaciones de cada sujeto respecto a la media estimada. Estos residuos son independientes y se distribuyen de forma normal con media cero y varianza común para todos los estudiantes, así la ganancia estimada se puede resumir como $G_i = (\beta_{10} + e_{1i}) - (\beta_{00} + e_{0i})$. Si dichas ganancias se miden como la influencia que las escuelas tienen sobre el cambio o crecimiento en el rendimiento académico de los estudiantes, pueden ser considerados como modelos de valor añadido (Navarro, 2013). En este caso el modelo de dos niveles se desarrollaría de la siguiente manera:

Ec. I.8

Nivel 1 (estudiantes):

$$Y_{tij} = \beta_{0ij} + \beta_{1ij} + r_{tij}$$

En el nivel de estudiante se estima la puntuación media del rendimiento de los estudiantes pertenecientes a la escuela *j* en el pre-test (β_{0ij}) y el post-test (β_{1ij}), ambos coeficientes son fijos. Como coeficientes aleatorios se encuentra r_{tij} que son residuos que se distribuyen de manera normal con medias cero y matriz de varianzas-covarianzas R (ecuación I.9). Con varianza σ^2_{r00} y σ^2_{r11}, respectivamente y covarianza σ^2_{r10}.

$$R = \begin{bmatrix} \sigma^2_{r00} \\ \sigma^2_{r10} \sigma^2_{r11} \end{bmatrix}$$ **Ec. I.9**

Cada coeficiente del nivel de estudiantes pasa a ser una ecuación del nivel de centro, de tal manera que el modelo en este nivel queda definido de la siguiente manera:

Ec. I.10

$$\beta_{0j} = \beta_0 + u_{0j}$$

$$\beta_{1j} = \beta_1 + u_{1j}$$

Donde β_0 y β_1 se refieren a la media general en el rendimiento inicial y la media general en el momento actual, respectivamente; u_{0j} es lo que aporta una determinada escuela j al estatus inicial y u_{1j} es el efecto diferencial del centro *j* sobre el post-test.

Así, se puede resumir la ganancia estimada en la siguiente ecuación:

$$G_j = (\beta_1 + u_{1j}) - (\beta_0 + u_{0j})$$ **Ec. I.11**

En estos modelos de ganancia estimada, el valor añadido alude al progreso de equidad que se obtiene al tener en cuenta que los factores que operan en el nivel actual, también operan en el nivel previo, de tal manera que el efecto de las covariables se considera tanto en el pre-test como en el post-test, quedando la medida definida de la siguiente manera $E(Y_{i1}|X_{i1}) - E(Y_{i0}|X_{i0})$ (Castro & Gaviria, 2009).

En definitiva, los modelos de ganancia han ido evolucionando hacia una metodología cada vez más precisa y fiable, pero cuentan con una gran debilidad y es que los modelos de ganancia tanto bruta como estimada no tienen en cuenta el efecto de los cursos perecederos a la hora de estimar la misma. Además, los modelos de ganancia residual pueden dar lugar a resultados sesgados al utilizar el rendimiento previo como covariable, debido a que los modelos de regresión asumen que las variables predictoras se miden sin error y, por tanto, las puntuaciones de los test se consideran fiables, sin embargo, al no considerar el error de medida a la hora de introducir el rendimiento previo puede ser problemático (Wiley, 2006; Choi et al., 2006).

En general, estos modelos al utilizar únicamente puntuaciones en dos ocasiones de medida, no aportan mucha información sobre el efecto que las diversas variables predictivas utilizadas pueden tener en la tasa de ganancia (Rogosa & Willet 1983). Por lo tanto, si lo que se desea es estudiar el progreso de un estudiante o de

un centro educativo a lo largo del tiempo es mucho más idóneo el uso de modelos de crecimiento o de valor añadido (Castro et al., 2009; McCaffrey et al., 2003a; Ponisciak & Bryk, 2005; Sanders & Horn, 1994; Singer & Willett, 2003; Stevens, 2006; Stevens & Zvoch, 2006; Zvoch & Stevens, 2003).

I.4.3. Modelos de crecimiento

En la actualidad existe un gran interés por el uso de modelos de crecimiento, como medio de mejora de la calidad de los sistemas de rendición de cuentas en el ámbito educativo. Esta tendencia creciente ha generado que muchas agencias estatales de educación, así como el personal y los responsables políticos, se enfrenten a tomar decisiones sobre cómo desarrollar e implementar un modelo de crecimiento en un sistema de rendición de cuentas en todo el estado. De esta manera, los modelos de crecimiento son vistos cada vez más como un enfoque estadísticamente válido en los sistemas de rendición de cuentas.

Los modelos de crecimiento pretenden medir el progreso del alumno mediante el seguimiento de sus puntuaciones en más de dos momentos temporales, con la intención de determinar si en promedio el estudiante avanza. Al comparar los datos de los mismos estudiantes a través del tiempo, el progreso puede definirse como el grado en que mejora el estudiante de acuerdo con el objetivo estatal o local establecido.

Estos modelos explican el progreso acumulativo del aprendizaje mediante la modelización de los logros de crecimiento en el tiempo. En este sentido, los modelos de crecimiento pueden variar, pero, en general, es de esperar que todos los estudiantes muestren algún progreso académico a lo largo de las mediciones, aunque algunas escuelas exhiben un mayor crecimiento, en promedio, que otras.

Dentro de los modelos de crecimiento existen dos grandes tipos, aquellos que consideran los datos anidados, donde generalmente los niveles de anidamiento son las ocasiones de medida, los estudiantes y los centros y se estiman mediante modelos multinivel o modelos jerárquicos lineales y los modelos lineales mixtos con medidas repetidas, que permiten que los alumnos varíen en las diferentes unidades de análisis, por ejemplo, que cambien de aula o escuela de un curso a otro. De esta

manera tradicionalmente, y tal y como se estudiará más adelante, los modelos de crecimiento han venido analizándose desde dos perspectivas:

1. **Modelos Jerárquicos lineales** (Bryk & Raudenbush, 2002), se caracterízan porque estima el estatus inicial y una pendiente de crecimiento, donde el valor añadido se refiere al residuo del nivel de aulas.
2. **Modelos lineales mixtos** (Sanders & Horn, 1994), donde se estima el promedio de logro en cada una de las mediciones realizadas. En este caso, el valor añadido se refiere a la ganancia que se produce entre dos ocasiones sucesivas.

Una ventaja de los modelos de crecimiento es que los estudiantes actúan como sus propios controles y, por lo tanto, representan las características de los mismos. Esto puede ser útil cuando hay cuestiones políticas sobre el establecimiento de objetivos de crecimiento, basados en los logros del pasado de los estudiantes con unas características específicas.

Una característica importante de estos modelos es que pueden no requerir información de entrada del estudiante (es decir, estado inicial), sin embargo, los modelos que operan con la intención de proporcionar proyecciones, se beneficiarán de los datos de entrada de los alumnos, como es por ejemplo el rendimiento inicial de los mismos.

Los modelos de crecimiento se aplican a los cinco objetivos de la investigación longitudinal (Baltes & Nesselroade, 1979):

1. La identificación directa del cambio intraindividual

2. La identificación directa de las diferencias interindividuales en el cambio intraindividual

3. El análisis de las interrelaciones en el cambio

4. El análisis de las causas (determinantes) del cambio intraindividual

5. El análisis de las causas (determinantes) de las diferencias interindividuales en intracambio individual

Según Wallis y Stepto (2007) los modelos de crecimiento pueden ser de diversa índole tal y como se puede observar en la siguiente tabla:

Tabla I.1. *Características de los modelos de crecimiento*

Modelo de Crecimiento	Características
Modelo de Mejora o Cambio de Cohorte	•Compara los resultados de una cohorte o clase en un grado determinado con las puntuaciones de una cohorte posterior de estudiantes en ese mismo grado. •El modelo en realidad compara dos grupos totalmente distintos de estudiantes, pero cuenta con la ventaja de que es un modelo fácil de implementar y entender. •Proporciona información de si los estudiantes de un grado en particular están consiguiendo las competencias establecidas de un año a otro. •Como aspecto negativo, el cambio que se produce en el porcentaje de alumnos que alcanzan el dominio puede ser debido a diferentes características de los estudiantes en cada cohorte, en lugar de un cambio en la eficacia escolar.
Modelo de Índice de Rendimiento	•Reconocen el crecimiento de los centros a través de una gama de niveles de logro académico acorde a un contexto educativo. No obliga a los centros a que consigan un determinado nivel de logro, sino que tiene en cuenta el crecimiento del centro de un año a otro, aunque este sea pequeño. •Son modelos utilizados por el estado, donde se da crédito a las escuelas si consiguen que los alumnos mejoren en su rendimiento. •No miden necesariamente el crecimiento académico de los estudiantes de manera individual, pero representan el cambio que se produce en el rendimiento de las escuelas de un año a otro.
Modelo de Crecimiento Simple	•Se refiere a la diferencia en las puntuaciones de la escuela de un año a otro. •Informan del crecimiento individual del estudiante de un curso a otro. •El crecimiento se calcula para cada estudiante de un año a otro y luego se promedia para calcular el crecimiento de la escuela.
Crecimiento de Modelo de Competencia	•Muestran si los estudiantes están en camino de cumplir con los estándares de rendimiento establecidos por el estado.

Fuente: elaboración propia a partir de Wallis y Stepto (2007)

Los modelos de crecimiento son similares a los modelos de valor añadido, pero estudian el cambio a nivel de los estudiantes y no a nivel de escuela. En realidad, los modelos de valor añadido son una variación de los modelos de crecimiento, ya que a partir de los mismos se estima un residuo que permite analizar el crecimiento a nivel de aula o del docente aislando su aportación de otros factores que son ajenos al entorno escolar.

En definitiva, los términos "valor añadido" y "modelo de crecimiento" son a menudo los métodos estadísticos utilizados para medir el crecimiento de la escuela y/o del estudiante con fines de rendición de cuentas. Los modelos de crecimiento se han centrado fundamentalmente en la rendición de cuentas de las escuelas de alto riesgo de fracaso escolar, pero algunos modelos de crecimiento, especialmente los modelos de valor añadido, se utilizan para evaluar a las escuelas o incluso maestros o programas de eficacia como herramienta para la mejora del sistema educativo.

La mayoría de los investigadores coinciden en que estas herramientas estadísticas, presentan un panorama más completo del rendimiento escolar frente a los modelos de estatus. Sin embargo, no están de acuerdo sobre el papel que los mismos deben desempeñar en la rendición de cuentas.

Las características que comparten los modelos de valor añadido que utilizan medidas de crecimiento a nivel de estudiantes son las siguientes (Navarro, 2013):

- Estudian el crecimiento académico de un mismo estudiante a lo largo de un periodo temporal determinado, utilizando más de dos ocasiones de medida. Lo importante es conocer el cambio que se produce de un momento a otro en un periodo concreto.
- El crecimiento se analiza generalmente, desde un enfoque multinivel donde se estima el estatus inicial (puntuación media en la primera ocasión de medida) y una pendiente de crecimiento (tasa de cambio en el rendimiento). También se puede analizar desde los modelos lineales mixtos, donde se incluye un coeficiente para cada ocasión de medida junto con el efecto de la escuela o docente asociado a dichos coeficientes.
- Con la finalidad de comparar las puntuaciones a lo largo del tiempo, las mismas deben encontrarse en una escala común, de esta manera se puede calcular la cantidad de crecimiento que los alumnos consiguen de una ocasión a otra. Para ello se suelen utilizar escalas verticales de rendimiento.
- El foco de interés en dichos modelos recae sobre la escuela o el profesor o incluso el distrito, pero no sobre el estudiante. Se pretende conocer qué cantidad de cambio del rendimiento del estudiante es debido a la escuela o al docente.

- Si se desea conocer si las escuelas rinden por encima o por debajo de lo esperado, es necesario conocer la cantidad de logro esperado, para ello es posible utilizar la media de crecimiento de los centros con similares características o, algún estándar fijado previamente.
- Dichos modelos pueden incorporar predictores contextuales a nivel de estudiantes o de escuelas.

El uso de variables contextuales como predictores en los modelos de valor añadido es un tema muy discutido, ya que son numerosos los autores que consideran que pueden afectar a las puntuaciones finales (Ballou, et al., 2004; Choi, et al., 2006; Ferrão, 2009; Haegeland & Kirkeboen, 2008; Hibpshman, 2004). Esta cuestión está relacionada especialmente con el uso de modelos de crecimiento, ya que muchos investigadores consideran que en estos modelos los sujetos ejercen su propio control debido a que se centran en el cambio y, por lo tanto, no requieren del uso de predictores de contexto (Stevens, 2005; Stevens & Zvoch, 2006). Sin embargo, otros autores si consideran que es importante la introducción de variables de contexto (Bryk & Raudenbush, 2002; Ferrão, 2009) para una mayor precisión en las estimaciones.

Aunque dichos modelos de crecimiento y valor añadido aportan información valiosa, hay que tener presente que si no se analizan con detalle distan mucho de ser la panacea, ya que cuentan con una serie de limitaciones que tienen que tenerse en cuenta a la hora de interpretar los resultados. Por ejemplo, el uso de distintas medidas de rendimiento a lo largo del tiempo se ven afectadas por el error de medición que se comete en cada una de ellas y, por ciertos factores de carácter causal. Estas circunstancias pueden dar lugar a una serie de fenómenos estadísticos, tales como el efecto de regresión a la media, egresión a la media, efecto techo, efecto suelo, etc. (Nesselroade, et al., 1980) que si no se tienen en cuenta pueden conllevar a interpretaciones sesgadas. Dicho de otra manera, las puntuaciones de ganancia que se obtienen en los modelos de crecimiento suele introducir ruido, ya que el aumento de las puntuaciones tiende a fluctuar, a pesar de que la verdadera capacidad de un estudiante normalmente no cambia mucho de un año a otro (Ballou 2002, Doran 2003). Esto sucede porque en cada prueba, el rendimiento de un estudiante es el resultado de su verdadera capacidad y las influencias aleatorias,

como distracciones durante la prueba, y el error de medición. Cuando las puntuaciones de las pruebas se restan entre sí, como en los modelos de crecimiento simples, se incrementan los errores de medición, de tal manera que el rendimiento "verdadero" se vuelve menos claro (Ballou, 2002).

Hay ajustes estadísticos para minimizar el "ruido", como la inclusión de las puntuaciones de otros sujetos y puntuaciones de rendimiento previo como covariable (Raudenbush, 2004). Otra forma es la creación de promedios móviles promediando el crecimiento durante varios años, para ofrecer una imagen más estable del rendimiento (Drury & Doran, 2003; Raudenbush 2004). Sin embargo, dichos ajustes se sumarán a la complejidad del modelo de crecimiento y pueden hacer que sea difícil explicar a los educadores por qué dos escuelas (o maestros), con ganancias de rendimiento similares, recibieron diferentes clasificaciones de efectividad (Ballou, 2002).

Un inconveniente importante propio de los diseños longitudinales, es que incluso el mejor sistema de recolección de datos no puede asegurar que todos los datos sean recopilados para cada estudiante, ya que los mismos, de un año a otro pueden cambiar de centro, o abandonan, o no asisten el día de la aplicación, etc.; lo que da lugar a que ciertos estudiantes tengan datos incompletos. Los datos faltantes pueden tener un gran impacto en los resultados del crecimiento, dependiendo de las características de los estudiantes que normalmente caen en ese grupo.

Una alta incidencia de la movilidad en una escuela produciría lagunas en los datos y podría distorsionar el "efecto" que la escuela o los profesores tienen sobre el rendimiento académico (McCaffrey, et al., 2003a). En este sentido no todos los modelos de crecimiento son capaces de incorporar información sobre los estudiantes que no tienen todas sus mediciones, aunque algunos lo hacen (CCSSO 2007), por ejemplo, el modelado mixto. Este tipo de modelos es el utilizado en el análisis de valor añadido de Tennesse (*Tennessee Value Added Assessment System*), donde mediante el uso de modelos lineales mixtos no excluyen a los estudiantes del cálculo simplemente porque les falte algún resultado en las pruebas, ya que cuentan con grandes series de datos. Además, los modelos mixtos permiten trabajar con valores perdidos, sin afectar a las estimaciones de los resultados cuando estos datos faltantes tienen una distribución aleatoria.

También hay algunos autores que consideran que una carencia de los modelos de crecimiento es que no están destinados a evaluar la tendencia creciente de la enseñanza en equipo. El argumento de crítica es el siguiente: si un estudiante está siendo enseñado por varios maestros, ¿cómo pueden dichos modelos aislar el efecto que cada docente tiene sobre el rendimiento del alumno? Nuevamente el modelo de Tennessee (que se abordará más adelante) es un ejemplo de superación de esta limitación, ya que es capaz de estimar los efectos de los maestros de manera independiente y aditiva; aunque varios profesores afecten al crecimiento académico del estudiante de manera simultánea (Kupermintz, 2003). Así pues, tal y como apuntan Wright, Sanders y Rivers (2006), los modelos de crecimiento pueden ser diseñados para dar cuenta de la enseñanza en equipo y la instrucción conjunta, aunque supone modelos muy difíciles de estimar y de interpretar.

A pesar de todas estas restricciones, la mayor limitación de dichos modelos es que requieren de un seguimiento en el crecimiento. Muchos estados, debido al coste que esto supone evalúan a los estudiantes en una sola ocasión de medida, no administran pruebas con cierto carácter temporal sobre un mismo individuo, y cuando el estado cuenta con un mayor número de pruebas, las mismas no están alineadas y carecen de un escalamiento, lo que imposibilita la comparación entre ellas (Yeagley 2007).

I.5. Modelos de valor añadido

En los Modelos de Valor Añadido (MVA, de aquí en adelante) se pretende conseguir los mismos objetivos que en los de ganancia y crecimiento, pero siguiendo otros procedimientos. Cuando los modelos de ganancia o de crecimiento pretenden aislar el efecto que las escuelas tienen sobre el logro de otros factores ajenos a su control, se denominan MVA.

Tal y como se ha comentado en este primer capítulo, los MVA son un tipo de modelo de crecimiento en el que se utilizan características generales de los estudiantes y /o logros y otros datos como controles estadísticos, con el fin de aislar los efectos específicos de una escuela, programa en particular o maestro sobre el progreso académico de los estudiantes.

Un MVA es un tipo de modelo de crecimiento, pero no todos los modelos de crecimiento son de valor añadido. En un modelo de crecimiento, las medidas de valor añadido estudian el cambio en el rendimiento individual de los estudiantes en un período de tiempo. Pero los MVA miden la cantidad en la que un recurso de instrucción en particular (maestro, programa de educación escolar, centro), contribuyen al cambio que se produce en el nivel de logro (Hershberg, Simón & Lea-Kruger, 2004). Esta es la distinción fundamental entre los MVA y los modelos de crecimiento, que a menudo se pierde en las discusiones acerca de la medición del crecimiento, ya que ambos términos se encuentran muy entrelazados.

Para Lissitz, Doran, Schafer y Willhoft (2006) los MVA son un tipo muy especial de modelos de crecimiento que añaden significado al cambio que se produce en el aprendizaje del estudiante, si se compara con la cantidad de cambio esperado o con los cambios que se han producido en otros estudiantes, con otras experiencias educativas (por ejemplo, diferentes profesores, escuelas o distritos) y, otros contextos (por ejemplo, nivel socioeconómico, género o etnia o rendimiento inicial). Es decir, el MVA es una herramienta estadística que facilita el aislamiento de los efectos de los docentes o las escuelas para poder probar su eficacia, independientemente de otros factores ajenos al centro, comparando los resultados con un valor de referencia.

Las distintas perspectivas del valor añadido se diferencian, principalmente, en cuestiones concretas relacionadas con el desarrollo de los modelos estadísticos utilizados para conseguir las estimaciones, aunque un MVA es típicamente la estadística compleja de todos los modelos de crecimiento. Sin embargo, si se usa correctamente, los MVA son posiblemente la herramienta estadística más poderosa para evaluar la eficacia de los maestros, las escuelas y los programas de educación.

I.5.1. Origen del análisis del valor añadido

El análisis del Valor Añadido (VA, de aquí en adelante) surge ante la necesidad de estudiar el cambio que se produce en los aprendizajes, junto con el movimiento de eficacia escolar, el estudio de los efectos escolares y la aproximación, desde una perspectiva más economista, de la función de producción educativa. El VA en educación, además de analizar el cambio en los aprendizajes, trata de estimar

puntuaciones que reflejen la aportación de las escuelas a ese cambio, independientemente de otros factores que no se encuentran bajo el control escolar.

Así, aunque en el ámbito educativo los MVA surgen inicialmente con la finalidad de dar respuesta a un tipo de sistemas de evaluación basados en la rendición de cuentas, es necesario señalar que el término VA proviene de sectores más económicos extendiéndose posteriormente al campo de la educación.

Los orígenes de dicho término se encuentran vinculados con aspectos relacionados con los bienes, la riqueza y la prosperidad (Saunders, 1999). En los contextos económicos, el término de VA cuantifica la diferencia existente entre lo que vale un producto y el coste que se invierte en la elaboración del mismo, es un proceso de entrada-producto (*input-output*). También dicho término alude a los beneficios que obtiene una empresa (outputs) quitando el precio de los costes (inputs) (Navarro, 2013).

La vinculación del VA al mundo de la educación se realiza mediante las aportaciones de investigadores procedentes del campo económico. Por ejemplo, Hanushek (1972) o Murnane (1975) estudiaron el efecto de los docentes y centros educativos sobre los resultados de logro, empleando funciones de producción educativa donde hablan de VA, refiriéndose al efecto que determinados factores educativos tienen sobre los resultados obtenidos en los sistemas educativos.

Estos autores junto con otros del mismo campo, trasladaron el análisis de costes-beneficios al ámbito educativo, considerándose el *Informe Coleman* de 1966 el germen de los estudios de la función educativa (Hanushek, 1971). Dicho informe dio lugar a estudios sobre eficacia escolar y la función de productividad educativa, cuyos resultados se relacionaban con variables controladas por las escuelas o administraciones (Navarro, 2013).

Una referencia temprana del uso del VA vinculado con las evaluaciones académicas en el estudio del crecimiento educativo se encuentra en Bryk y Weisberg (1976), quienes modifican la concepción de *input-output* utilizando modelos más innovadores, diferentes a la regresión múltiple o análisis de covarianza. Estos nuevos modelos son más dinámicos, ya que consideran que un determinado programa educativo puede afectar al logro de los estudiantes introduciendo cambios en la tasa de crecimiento de los mismos, es decir, diferentes factores

escolares, tales como programas, docentes, centros, etc., pueden producir cambios en los resultados de logro.

Dichos autores introducen en el campo educativo una concepción de VA donde tiene en cuenta dos puntuaciones de logro, una, antes del tratamiento (pre-test) y otra, después (post-test). A diferencia de los economistas, en vez de estimar una función lineal sobre la ganancia bruta se basan en la utilización del pre-test y del post-test, como variable de respuesta relacionándolas con la edad, ya que pretenden diferenciar la ganancia que se produce por maduración de la que se produce por el efecto de un determinado programa.

Esta nueva concepción de VA pretende estudiar la diferencia entre los valores observados, con los predichos en el pre-test y post-test, considerando el VA como la diferencia entre la puntuación observada tras la aplicación del tratamiento (Navarro, 2013). Según este mismo autor, la importancia de dichos modelos reside en entender el VA como la diferencia entre lo que se espera por desarrollo natural y lo que se observa. Estos modelos pueden ser considerados como un paso anterior a los actuales Modelos Jerárquicos Lineales que se utilizan para estimar el VA.

En esta primera concepción de VA en el campo educativo, el modelo se formula matemáticamente de la siguiente manera:

$$Y_i = \beta a_i + \varepsilon_i \qquad \textbf{Ec. I.12}$$

Donde Y es la puntuación de logro del sujeto *i*; a_i se refiere a la edad o la medida de tiempo que se utilice (años, meses, ocasiones de medida), β es el crecimiento que se considera que es conocido y constante y, por último, ε_i es el residuo aleatorio que puede cambiar entre las distintas ocasiones de medida y de una ocasión a otra.

Aunque fueron Bryk y Weisberg (1976) quienes introdujeron el concepto de VA en el campo educativo, la aparición de dicho fenómeno en este contexto se debe a la supeditación de tres factores fundamentales: la medición de la eficacia escolar, la rendición de cuentas y el desarrollo de complejos modelos estadísticos (Martínez-Arias et al., 2009). A continuación, se va a estudiar cómo cada uno de estos factores influyeron en el desarrollo del VA en el campo educativo.

A) La investigación sobre la efectividad de las escuelas

Dentro de la investigación sobre la eficacia de las escuelas, fue a través del *Informe Coleman* (1996), mencionado con anterioridad, cuando se empezó a estudiar la relación existente entre el logro académico y variables familiares y contextuales. Tras dicho informe se realizaron estudios posteriores, a través de los cuales se hacían comparaciones entre las escuelas, con el fin de determinar la calidad de los aprendizajes e introducir mejoras en aquellos sistemas que se consideraban menos eficaces.

Realmente los orígenes del estudio de la Eficacia Escolar se encuentran fundamentalmente en Reino Unido y Estados Unidos, donde numerosos estudios realizados en la década de los sesenta y setenta mostraban que los factores que más afectaban al logro eran aspectos tales como la étnia, raza, estatus socioeconómico, etc., mientras que las escuelas parecían no afectar mucho a los resultados obtenidos. (Coleman et al., 1966).

Estos estudios contaban con un gran número de limitaciones metodológicas que la investigación posterior suplió. Aun así, demostraron que, aunque los factores socioeconómicos y culturales tenían una gran influencia en el nivel de logro de los estudiantes, las escuelas presentaban efectos significativos sobre el rendimiento de los alumnos.

En este contexto se empiezan a desarrollar los análisis de valor añadido con la finalidad de estudiar, cómo afectan las escuelas en el nivel de logro de los estudiantes que asisten a las mismas, aislando la influencia de otros factores relacionados con el aprendizaje de los alumnos, como puede ser su estatus socioeconómico.

La metodología utilizada en el estudio de la eficacia era fundamentalmente cuantitativa, basada en test estandarizados y análisis complejos que separaban los resultados de rendimiento de los estudiantes, de otros factores ajenos al campo educativo.

En un primer momento, se estudió lo que se puede denominar como modelos de estatus, ya que se trataban de estudios transversales realizados en un único momento temporal, algo que resultaba y resulta insuficiente para el estudio de los

aprendizajes (Willms & Raudenbush, 1989). Sin embargo, la investigación más temprana ha modificado estos métodos mediante diseños más sofisticados, donde se estudia el progreso de los aprendizajes en un mismo estudiante a lo largo del tiempo, dejando de lado las evaluaciones de estatus y centrándose en los diseños longitudinales.

Estos nuevos estudios longitudinales realizados por diversos autores (por ejemplo, MacBeath & Mortimore, 2001; Reynolds & Creemers, 1990; Scheerens, 2005; Teddlie & Reynolds, 2000) han mostrado que las escuelas afectan de manera significativa en los niveles de logro, haciendo que estos sean diferentes. Asimismo, en estos modelos se ha demostrado que las variables inputs (recursos de los países desarrollados), a excepción de variables relacionadas con la experiencia y formación docente, no influyen tanto en el logro (Hanushek, 2003), como demostraban los modelos iniciales.

Estos datos han hecho que los inputs pierdan importancia y que, las investigaciones se centren más en los resultados de las evaluaciones educativas, mediante el uso de medidas de VA que proporcionan información precisa sobre el desempeño escolar, promoviendo de esta manera, estrategias que mejoren entornos de desempeño ineficaz.

B) Rendición de cuentas

La aparición del concepto de VA en el campo educativo, está íntimamente vinculado a la adopción de sistemas escolares públicos, basados en la rendición de cuentas.

El objetivo de los sistemas de rendición de cuentas, tal y como se ha estudiado en el apartado I.2, es la comparación del uso que hacen las diversas instituciones públicas con la financiación que reciben. Los sistemas de rendición de cuentas pretenden informar sobre los resultados obtenidos en las evaluaciones realizadas, con la finalidad de buscar buenas prácticas que mejoren dichos resultados o con la intención de proporcionar a los centros incentivos o sanciones acorde a los logros obtenidos. En cualquier caso, independientemente del fin, la rendición de cuentas debe ser entendida como un medio para la consecución de una serie de metas educativas.

Aunque el movimiento de la rendición de cuentas tiene su máximo apogeo con la ley NCBL, ya años anteriores existía un interés creciente dentro de las políticas generales, especialmente en Estados Unidos y otros países, por la evaluación externa de las escuelas como medio de supervisión y mejora continua de los sistemas educativos (Hanushek & Raymond, 2004; Kane & Staiger, 2002; Taylor & Nguyen, 2006).

Los primeros sistemas de rendición de cuentas se centraban en medir el número de recursos y procesos disponibles, por ejemplo, la ratio de las aulas, las credenciales docentes o el número de días de docencia. Sin embargo, actualmente se centran en el estudio de los resultados (Hanushek & Raymond, 2004), esto conlleva la necesidad de una metodología potente que permita establecer diferencias entre las escuelas. En este contexto los MVA han adquirido un papel muy importante dentro de los sistemas de rendición de cuentas.

En Martínez-Arias (2009) se pueden encontrar diferentes tipos de sistemas de rendición de cuentas basados en los resultados, por ejemplo, el uso de MVA para identificar casos extremos, es decir, niveles de logro muy altos o bien muy bajos, MVA que evalúan el efecto del profesorado, de un programa o de la escuela, MVA para el establecimiento de ranking, etc. En cualquier caso, estos modelos tienen que usarse con cautela cuando se pretenden usar como un medio para tomar decisiones de carácter sancionador sobre los profesores o escuelas (Braun, 2005; McCaffrey et al., 2003a; Raudenbush, 2004), ya que los resultados obtenidos mediante la aplicación de los mismos pueden verse afectados por determinados artefactos estadísticos (por ejemplo, efecto de regresión a la media) que pueden sesgar los resultados sino se desconoce su comportamiento.

Algo muy común en los sistemas de rendición de cuentas es la publicación de los resultados obtenidos en las evaluaciones educativas, con la finalidad de dar a conocer los puestos que ocupan cada uno de los centros educativos. Esto proporciona información a las familias sobre el rendimiento de las diferentes escuelas, lo que les puede servir de ayuda en la toma de decisión de la elección de la misma, para la educación de sus hijos.

En este sentido, hay partidarios a favor de la divulgación de dicha información, como medio para mejorar el nivel de logro de los estudiantes (Glenn &

de Groof, 2005). Por ejemplo, en una investigación realizada por Fuchs y Wossman (2007) mediante el análisis de datos procedentes de evaluaciones internacionales, descubrieron que diversas formas de rendición de cuentas, junto con la autonomía y la elección de escuelas, se vinculaban con niveles de logro altos de los estudiantes.

Sin embargo, hay otros autores que consideran que hay que tener en cuenta que estas comparaciones entre escuelas, solo son posibles dentro de un marco contextualizado, donde los centros compartan características afines (Ballou, 2002).

Lo importante a la hora de proporcionar dicha información procedente de los sistemas de rendición de cuentas, es que cuando se publiquen los resultados no se tomen represalias hacia los centros o docentes, sino que se use con fines de mejora. La comparación de dichos resultados entre escuelas es una tarea compleja, que ni el uso de sofisticados MVA pueden garantizar la completa equidad de los resultados. Lo que sí es cierto, es que dichos modelos son más potentes que los modelos de estatus, ya que estos últimos tan solo permiten conocer el estado de los sistemas educativos en un momento determinado.

C) Modelos estadísticos

A pesar de la polémica generada por la ley NCBL sobre la medición de la eficacia dentro de los sistemas educativos, su enfoque en la rendición de cuentas de las escuelas, supuso un impulso muy importante de los modelos estadísticos de medición, incluso sobre los modelos de valor añadido.

Los MVA en la evaluación no se hubieran desarrollado sin modelos estadísticos, que permitieran descomponer los cambios en los aprendizajes de los estudiantes en diferentes fuentes de variación, analizar variables vinculadas con los distintos niveles y, tener en cuenta que los datos de las escuelas no son independientes (Martínez- Arias et al., 2009).

El uso de los primeros MVA, donde las puntuaciones brutas de las pruebas de los estudiantes servían para el establecimiento de rankings entre centros, correlacionaban con el estatus socioeconómico de los mismos, lo que hizo que dicha metodología no fuera considerada la más adecuada para la clasificación de las escuelas (McCall, Kingsbury & Olson, 2004). Con el propósito de poner fin a dicha problemática se empezaron a realizar análisis multivariados de corte transversal

(Bethell, 2005). En estos análisis, las puntuaciones medias de las pruebas de los centros, incluyen un análisis de regresión unido a una serie de variables contextuales de los estudiantes. El objetivo era ordenar a las escuelas de acuerdo a sus residuos de regresión, ya que los mismos son considerados como efectos de la escuela. Los centros que tenían residuos negativos eran calificados como problemáticos mientras que, los que tenían unos residuos positivos eran considerados como escuelas exitosas.

Aunque estos últimos modelos eran más eficaces para la estimación del VA que las puntuaciones brutas, el problema de muchos de ellos, tal y como ocurrió con los trabajos iniciales de Coleman et al. (1966), es que no tenían en cuenta el supuesto de independencia de los errores aleatorios, ni la estructura jerárquica de los datos educativos, por la cual los estudiantes están anidados en las escuelas y estas a su vez se encuentran anidadas en distritos, lo que violaba el supuesto de independencia, ya que los niveles se encuentran relacionados.

Ya Thorndike en el año 1939 hizo notar, lo que más tarde Robinson (1950) catalogó como «falacia ecológica»[3] , es decir, los problemas que suponía utilizar estimaciones procedentes de la correlación y regresión del nivel de centro, para hacer inferencias sobre subgrupos.

Todo esto hizo que en los ochenta empezaran a aparecer, en áreas ajenas a la educación, los primeros modelos conocidos como modelos multinivel, modelos jerárquicos lineales o modelos lineales mixtos (Goldstein, 1987; Raudenbush & Bryk, 1986). Estos se caracterizaron porque permitieron la división de la varianza en distintos niveles y, la introducción de diversos predictores, siendo posible el estudio de la interacción entre los mismos.

Dicha metodología, mucho más sofisticada, permitía poner fin a un gran número de sesgos que afectaban a las inferencias estadísticas, por ejemplo, errores tipo I, heterogeneidad, errores de agregación, etc. Y además posibilitaba disgregar el efecto de los centros de otros factores relacionados con los alumnos, dando lugar a modelos de logro más contextualizados.

[3] Para una explicación más detallada de la falacia ecológica consultar Burnstein (1980)

Aunque estos modelos suponen un gran hito en la medición del VA, contaban con el problema de la inestabilidad temporal de los resultados, de tal manera que centros con un nivel de logro alto en un año, podían pasar a un nivel mediocre al año siguiente.

La extensión de estos modelos al estudio de datos longitudinales (Wilms & Raudenbush, 1989), supuso un gran avance para estudiar las trayectorias del crecimiento y ver cómo afectan las escuelas sobre el mismo (Martínez et al., 2009).

En la actualidad estos modelos han adquirido cada vez una mayor fuerza, siendo su uso muy común en un gran número de evaluaciones de los países miembros de la OECD. Fitz-Gibbon (1997) realizó un análisis de dichos modelos como método de evaluación y concluyó, que un modelo de estas características podía ser considerado como la base de un sistema nacional de valor añadido altamente eficaz y válido; ya que, mediante varios ajustes, los datos sobre el crecimiento de los estudiantes se pueden convertir en indicadores de valor añadido en el ámbito escolar.

Otros autores que también han trabajado sobre dichos modelos, por ejemplo, Gaviria y Castro, 2005; Gelman y Hill, 2006; Raudenbush y Bryk, 2002, están de acuerdo en considerar, que dichos modelos no se hubieran podido aplicar sin un potente software estadístico que permita complejos procesos de estimación. En la actualidad, para el análisis de los mismos se cuentan con programas estadísticos tales como, MLWIN, HLM, SPSS, SAS, MPLUS, STATA, SYSTA, R, entre otros.

I.5.2 Definición de valor añadido

Como se ha comentado en este capítulo, el término de Valor Añadido procede de contextos puramente económicos, en este sentido Spours y Hodgson (1996) señalan que el valor añadido se utiliza como término ambiguo para describir un campo inespecífico compuesto de diversas prácticas y proyectos. Dichos autores consideran que el término de valor añadido se origina en el sector de la economía, cuando se hace referencia a la diferencia entre los insumos, como las materias primas y la energía, y los productos finales que están representados por el valor de las ventas. Para Spours y Hodgson (1996) el concepto económico de valor añadido ha adquirido un significado adicional cuando se asocia a la necesidad de que las economías occidentales avanzadas, se centren en la fabricación de manufacturas de

alta calidad como una forma de competir exitosamente con las economías que crecen rápidamente. En este sentido, el término valor añadido se aplica a los bienes cuya calidad y valor se incrementan por altos niveles de tecnología y habilidad, en el proceso de fabricación.

En el ámbito educativo no es fácil encontrar las primeras definiciones estables del concepto de VA. Se cree que a principios de los años noventa en Reino Unido, personas de círculos políticos hablaban con distintos grados de conocimiento acerca de la necesidad de comprender y analizar el valor añadido en educación (Saunders, 1999).

Fue en el año 1990, cuando el término de VA se consideró de uso bastante común, como para aparecer en el título de un documento clave, publicado por el Consejo de Premios Académicos Nacionales, denominado *The measurement of value added in higher education*[4] (La medición del valor añadido en la educación superior). En dicha obra, el valor añadido era considerado una medida de calidad, que reflejaba como la experiencia educativa aumentaba el conocimiento, las habilidades y destrezas de los estudiantes (McGeevor, Giles, Little & Brennan, 1990). Se consideraba que una institución de alta calidad sería aquella que mejorara en gran medida el logro de sus estudiantes.

En el año 1992, la Comisión Nacional de Educación consideró que el término valor añadido era lo suficientemente bien comprendido, incluso por personas ajenas a la profesión educativa, como para ser implantado en el estudio de las escuelas, lo que llevo al inicio de los primeros intentos de medida del VA en el campo educativo (McPherson, 1992).

Como término explícito, su primera aparición en las escuelas parece haber sido en el sector de la educación superior, concretamente en Educación Secundaria Obligatoria (ESO), dentro del contexto del desarrollo de indicadores de desempeño institucional y programas de investigación sobre la efectividad escolar. Este sector de educación superior se interesó particularmente por evaluar la eficiencia y la eficacia de las instituciones educativas, con el fin de justificar las asignaciones de fondos. Sin embargo, el resto de sectores escolares estaban más preocupados por

[4] Para más información revisar McGeevor, Giles, Little y Brennan, 1990.

hacer comparaciones más justas entre las instituciones educativas, ya que pretendían explicar la aparente disparidad en la tasa de rendimiento, para niveles de financiación similares, en diferentes instituciones o autoridades educativas locales.

A pesar de las distintas conceptualizaciones del VA, todas tienen en común la misma pregunta de fondo: "¿Cómo se puede medir el progreso del alumno/alumna de tal manera que arroje luz sobre el rendimiento de las instituciones? Esta pregunta es probablemente la clave para entender los principios y exigencias metodológicas del valor añadido, que se estudiará más adelante.

Las políticas educativas del año 1998, en el sector de la enseñanza superior, entendían el valor añadido como una medida del rendimiento de los estudiantes, que tiene en cuenta los efectos de los insumos diferenciales de los alumnos (es decir, las calificaciones de entrada) (McGeevor et al., 1990). En esta concepción los logros estudiantiles, se utilizaban para evaluar la calidad institucional de las funciones de enseñanza y aprendizaje.

Dicha definición muestra la fuerte conexión que había y que se mantiene hasta hoy, entre la eficiencia y calidad y la preocupación por la financiación. Por lo tanto, el valor añadido educativo comenzó como una verdadera preocupación económica.

En este contexto, la cuestión de la calidad se relacionaba con la aportación que la institución educativa o servicio hacia sobre los participantes. El valor añadido era la mejora que se producía en el estudiante por la contribución del curso.

La definición de VA no ha evolucionado mucho hasta nuestros días, haciendo que no exista consenso común sobre la misma, lo que ha conllevado a que no haya una única definición. Pero, se puede distinguir, entre definiciones de carácter más teórico y otras que dependen más de la forma en que el valor añadido se operacionaliza.

A nivel teórico se puede destacar la definición de VA de Bennett (2001), quien dice que cuando se habla de VA se hace referencia a lo que se ha mejorado sobre las capacidades o conocimientos de los estudiantes, como consecuencia de su educación en un colegio o universidad en particular. El valor añadido es la diferencia entre los

logros de los alumnos cuando han completado su educación y, lo que ya habían alcanzado en el momento en que comenzaron.

Harvey (2002), señaló que el valor añadido se refiere a la mejora de los conocimientos, habilidades y capacidades de los alumnos y la potenciación de ellos, como estudiantes críticos y reflexivos a lo largo de toda la vida. Dicho autor considera que el VA se ha convertido en un foco de atención, como resultado de las tablas de clasificación (rankings) publicadas en todos los niveles de la educación y también, debido al creciente interés en la medición de la empleabilidad. Sin embargo, a pesar de ello es difícil evaluar el valor añadido, ya que la mayoría de los intentos, se han basado en la medición de entrada y salida de los grados o habilidades, utilizando indicadores poco precisos.

El análisis cuantitativo del valor añadido es difícil por una gran variedad de razones, entre ellas, el establecimiento de puntos de referencia, problemas de medición y la atribución del valor añadido a programas, en lugar de algún otro factor. A pesar de ello, la evaluación del valor añadido debe estar orientada a la mejora, a la buena relación calidad-precio y, al enfoque transformador de la evaluación de la calidad a nivel de programas (Harvey, 2002).

En el año 2004, Tekwe et al. utilizan dicho término para referirse a métodos de evaluación de los niveles de logro de los profesores y las escuelas, que se centran en medir la ganancia que se produce de un año a otro en el nivel de rendimiento de los estudiantes.

Otra definición, altamente conocida, es la aportada por Martínez-Arias et al., (2009) quienes consideran que el valor añadido es la "contribución de la escuela al progreso neto de los estudiantes hacia objetivos de aprendizaje establecidos, una vez eliminada la influencia de otros factores ajenos a la escuela que pueden contribuir a dicho progreso" (p.17).

Braun, Chudowsky y Koening (2010) mencionan que, en el contexto educativo, la metodología del Valor Añadido hace referencia a los esfuerzos para medir los efectos, que tienen en el rendimiento de los estudiantes sus actuales profesores, escuelas o programas educativos, teniendo en cuenta las diferencias en rendimiento previo y otras medidas de características contextuales que los estudiantes llevan consigo a las escuelas.

Según la OECD (2011b) el valor añadido alude a la contribución del centro al nivel de logro de los estudiantes entre al menos dos medidas, una de ellas referidas al logro previo (pre-test) y otra al logro actual (post-test), la diferencia entre ambas se usa para la toma de decisiones de índole político-educativo.

Más recientemente Thieme, Prior, Tortosa-Ausinac y Gempp (2016) consideran el valor añadido como la contribución que la escuela realiza al nivel de logro del estudiante, controlando el rendimiento previo de los mismos.

En esta tesis el VA es considerado como la verdadera aportación que las escuelas realizan sobre los aprendizajes de los estudiantes, aislando otros factores que pueden influir en el crecimiento de los mismos. En este sentido hay que tener en cuenta, tal y como se demostró en el informe Coleman, que los resultados de las puntuaciones obtenidas por los estudiantes se encuentran influenciadas por el contexto familiar, experiencias y procesos educativos, motivación, etc. Por lo tanto, a la hora de evaluar el impacto de una escuela concreta hay que tener presente estos aspectos, ya que los resultados brutos del nivel de logro no solo se ven afectados por la acción escolar.

Una medida que está contaminada por factores no vinculados al centro, no es conveniente utilizarla para realizar comparaciones. Asimismo, tampoco se recomienda comparar dos cursos distintos, con el fin de estudiar qué escuela está aportando más ganancia al aprendizaje de los estudiantes, ya que dicha ganancia puede verse afectada por factores ajenos a la escuela.

Para conocer la aportación de los centros, la ganancia, se necesita un criterio que determine si una escuela produce VA en los estudiantes. Generalmente el VA se estima una vez que se tienen en cuenta las diferencias iniciales existentes entre los estudiantes, como la diferencia entre la ganancia observada y esperada (Goldschmidt et al., 2005).

Aunque todas las definiciones propuestas se centran en definir el VA, como la mejora que se producen en el nivel de logro de los estudiantes gracias a la contribución, principalmente, de la escuela, un aspecto diferenciador de los MVL es la introducción de distintas medidas de logro, que permiten una estimación más precisa de lo que aporta la escuela, en el nivel de rendimiento de los estudiantes

Doran e Izumi (2004) mostraron las ventajas de los MVA que utilizan varias mediciones a lo largo del tiempo, en comparación con los modelos de estatus que ofrecen una imagen estática del logro escolar. Sin embargo, a pesar de la gran información que aportan los modelos con más mediciones, la mayoría de los sistemas educativos se centran en realizar evaluaciones de tipo transversal (debido a que supone menos coste) donde no se siguen a los mismos sujetos a lo largo del tiempo, sino que generalmente se comparan distintas cohortes pertenecientes a años distintos. Esto no significa que no se pueda estimar el VA de unos años a otros, sino que hay que ser muy comedidos con las decisiones que se toman a partir de estos modelos, ya que, en los mismos se pierde información y se encuentran contaminados por el efecto de regresión a la media, que como se estudiará posteriormente, es más difícil de paliar que en los modelos longitudinales en los cuales se sigue a los mismos estudiantes en más de dos ocasiones de medida.

En este sentido, Willett (1989; 1994) hizo una distinción entre los modelos de ganancia basados en dos únicas mediciones y, los modelos de crecimiento que usan tres o más aplicaciones, demostrando que estos últimos modelos proporcionan, un análisis más detallado de la contribución que la escuela hace a la mejora del desempeño académico.

Así, diferentes formas de medir el cambio en los aprendizajes: como estatus, ganancia o crecimiento, da lugar a diferentes formas de operacionalizar el valor añadido.

Teniendo en cuenta el carácter práctico del VA y los análisis realizados sobre el mismo, se han definido distintos modelos. Murphy (2012) considera que los modelos que se discuten más comúnmente dentro de la literatura de investigación, pueden catalogarse bajo tres clasificaciones amplias: modelos de puntuaciones de ganancia, modelos de puntuaciones de ganancia residual y modelos longitudinales multivariados. Hay ventajas y desventajas asociadas con cada tipo de modelo, y ninguno de ellos ha sido considerado como idóneo en todas las condiciones. La siguiente tabla presenta las características definitorias, ventajas y desventajas, y ejemplos prototípicos de cada uno de ellos

Tabla I.2. *Clasificaciones y ejemplos de modelos de valor añadido con ventajas y desventajas asociadas*

Modelo	Características	Ventajas	Desventajas	Ejemplo
Puntuación de ganancia	•Se comparan los resultados de las pruebas de un año para otro. •Se comparan los beneficios a través de los profesores. • Asume que todas las puntuaciones están en la misma escala. • Las puntuaciones de los sujetos son transversales.	• Intuitivo para las partes interesadas. • El crecimiento del estudiante es la variable de interés. •Modelos simples que son fáciles de implementar y entender. • Cuentan los factores que afectan sólo al nivel de logro de los estudiantes.	• Ignora la información sobre los estudiantes. • Los resultados de las pruebas deben estar escaladas verticalmente. •Los estudiantes con datos perdidos están excluidos del modelo.	• Modelos de puntuación de ganancia simple. • Puntuaciones de ganancia transversal.
Puntuación de ganancia residual	• Las puntuaciones actuales son regresadas a las puntuaciones previas y, los residuos son considerados como desviaciones del crecimiento esperado. •A menudo incorporan características del estudiante al modelo. •Las puntuaciones de los sujetos son transversales.	•Se puede hablar de predicción de logro, aunque las puntuaciones no estén escaladas. •Fácil de especificar y de ajustar. •Se pueden especificar como modelos no lineales.	• Se ajustan modelos por separados en cada año, por lo tanto, los datos evaden información previa como, por ejemplo, la influencia del maestro. • Los estudiantes con datos perdidos son obviados del modelo.	• Modelos de ajuste de covarianza. • Modelo de rendición de cuentas como el modelo de Valor Añadido de Dallas.
Modelos Longitudinales Multivariantes	• Modelan las distribuciones de los resultados de los estudiantes. • Son modelos longitudinales.	• No se ven afectados por los valores perdidos. • Utilizan toda la información del estudiante.	• Modelos difíciles de estimar. • Requieren de software potentes.	• Evaluación basadas en Modelos de Valor Añadido. • Modelo de efectos cruzados aleatorios.

Fuente: elaboración propia a partir de Murphy (2012)

En definitiva, el VA se ha venido utilizando principalmente como una medida de calidad de los centros, a través de la cual se pretende realizar un estudio de la situación actual del sistema educativo. Estos análisis suponen una serie de beneficios y/o utilidades importantes en el campo educativo, pero también, cuentan con importantes dificultades. Ambos aspectos serán estudiados a continuación.

I.5.3. Utilidades y beneficios del análisis del valor añadido

Diversos son los factores que han desencadenado el desarrollo de los MVA, entre los más destacados se puede considerar el interés por el estudio de la eficacia de las escuelas, la implantación de políticas educativas basadas en la rendición de cuentas y el desarrollo de técnicas estadísticas avanzadas.

Los modelos de valor añadido suponen un progreso respecto a otro tipo de medidas de logro, que actualmente son las más utilizadas en las evaluaciones educativas, por ejemplo, las medidas de estatus. Los MVA aportan información muy relevante, al hacer uso de variables tales como, el rendimiento previo o características contextuales de los alumnos, que permiten un análisis más refinado del crecimiento académico de los estudiantes y, de otros factores que pueden influir en el nivel de logro de los mismos. Todo esto ha conllevado a que dichos modelos, sean considerados de gran utilidad para aquellos países que están en proceso de mejora e innovación educativa (OECD, 2011b). La utilidad y beneficios de los MVA residen en que:

- Permiten identificar aspectos, que influyen en menor o mayor medida, en los procesos de aprendizaje de todos los niveles educativos.
- A partir de los resultados obtenidos en los mismos, se pueden introducir una serie de factores que mejoren los procesos de aprendizaje.
- Son capaces de esclarecer los sistemas de rendición de cuentas, haciendo que sean procesos más equitativos, que busquen la mejora de los sistemas educativos dentro de un marco contextualizado.
- Desarrollan sistemas de información que permiten a las escuelas analizar y evaluar su desempeño y fortalecer el sistema de evaluación.
- Mejoran la asignación de los recursos materiales, económicos, humanos, etc., ya que tienen en cuenta las diferencias socioeconómicas de los alumnos.

Además, las evaluaciones que se basan en el valor añadido suponen una serie de ventajas, tanto para los responsables de los sistemas educativos, como para los profesores. Para los responsables de los sistemas educativos, dichas evaluaciones les ayudan a tomar decisiones más afines al contexto en el cual se desenvuelve la acción educativa, y a los profesores les permite realizar evaluaciones más ajustadas a los resultados obtenidos en las evaluaciones de valor añadido.

El valor añadido proporciona información muy relevante para el diseño y desarrollo de programas, que mejoren el estado educativo de las escuelas (OECD, 2011b). Todo ello desde una perspectiva contextualizada, donde se identifican aquellos factores que más afectan al rendimiento de los estudiantes y, por ende, de las escuelas.

Generalmente, los resultados obtenidos mediante la estimación de los MVA se orientan fundamentalmente a tres ámbitos.

A) Mejora de los sistemas educativos

Los modelos de valor añadido como medio para la mejora escolar, se benefician del análisis de los datos procedentes de las características específicas de los estudiantes en una escuela concreta. Estas variables proporcionan resultados más ajustados, que permiten a las escuelas adoptar medidas para la mejora del centro y de la labor docente, ya que informan sobre aquellos factores más débiles y fuertes del desarrollo académico de los estudiantes.

Con los resultados obtenidos en los MVA los docentes pueden formular objetivos que mejoren las metodologías utilizadas, adaptando las mismas a las necesidades de aprendizaje de sus estudiantes. En este sentido, los modelos de valor añadido ayudan a realizar un seguimiento más adecuado de aquellos estudiantes que necesitan mayor apoyo y, permiten al profesorado un mayor desarrollo de su labor como profesionales.

Los MVA como medio de mejora escolar, aportan una serie de diseños innovadores, donde a partir de la situación de partida se puede diseñar una metodología adaptada a las necesidades educativas, que posteriormente puede ser evaluada con la finalidad de estudiar la influencia que las mismas han tenido sobre el nivel de logro de los estudiantes.

Una de las aportaciones más importantes del VA en esta área, es que permite reforzar los logros realizados por la escuela, ya que se puede conocer lo que realmente aporta el centro al desarrollo académico de los estudiantes. Esto permite a los centros reflexionar sobre su propia práctica y establecer las medidas necesarias para la mejora educativa. En este sentido, el VA tiene una función diagnóstica de la situación de los centros, que es aplicable a cualquier sistema de evaluación educativa, y una función de optimización, ya que los análisis realizados permiten la introducción de una serie de cambios que mejoren los resultados.

Desde esta perspectiva de mejora, el VA es capaz de definir perfiles sobre la trayectoria académica de los estudiantes, de tal manera que es posible conocer que estudiantes progresan de manera adecuada, y aquellos que se encuentran más atrasados (Doran & Izumi, 2004). Con estos resultados se pueden diseñar o destinar una serie de recursos o programas, para la mejora del nivel de logro de estos alumnos.

Las combinaciones de las trayectorias del progreso académico de los estudiantes, junto con el VA de las escuelas permite realizar proyecciones futuras sobre el rendimiento académico (OECD, 2011b). Esto supone a las escuelas poder tener una perspectiva futura de aquellos estudiantes, que podrán conseguir los objetivos establecidos. Mediante la comparación de los resultados obtenidos con los estándares establecidos previamente o bien, mediante la comparación de esos resultados con otro grupo de estudiantes de características similares.

Estas proyecciones futuras sobre el desarrollo de los niveles de logro de los estudiantes y en definitiva de las escuelas, pueden ayudar a la planificación de programas de mejora, distribución de recursos, toma de decisiones, etc., de tal manera que se puede actuar con antelación sobre los posibles problemas futuros.

Para que se pueda aprovechar al máximo los beneficios del VA, es necesario que toda la información que aporta, sea utilizada por los agentes que influyen en los procesos y resultados de logro. Si dichos agentes (profesores, directores, inspectores, etc.) comprenden dicha información tendrán una ayuda extra para la mejora y organización de los sistemas educativos.

No se puede finalizar dicho apartado sin hablar del modelo de valor añadido de *Tennessee Value-Added Assessment System* (TVAAS) (Sanders & Horn, 1994; Sanders, et al., 1997), ya que fue el primer modelo de valor añadido utilizado para la mejora escolar.

Actualmente dicho modelo recibe el nombre de EVAAS (*Evaluation Value Added Assessment System*), y es desarrollado por la empresa de software estadístico SAS[5] mediante análisis de Modelos Lineales Mixtos, ya que cuentan con datos de carácter longitudinal, y una estructura anidada o cruzada.

Este modelo, EVAAS, surge debido a un trabajo realizado en la Universidad de Tennesse por William Sanders donde, debido a una ley estatal implantada en 1993, se exigía a todas las escuelas de los distintos distritos que recogieran datos de los estudiantes y se los proporcionaran a Sanders. Esta ley prohibía el uso de los análisis de VA como medio para la rendición de cuentas de los docentes y escuelas, de tal manera, que solo se podían utilizar como medio de mejora para el desarrollo escolar, dejando a los distritos la decisión de utilizar o no, los resultados de dicho modelo.

Los informes de valor añadido EVAAS comenzaron con informes a nivel de distrito en 1993 y se ampliaron a los informes escolares en 1994, y a los informes a nivel de maestros en 1996. El término Valor Añadido en esta evaluación, se refiere a un análisis estadístico utilizado para medir el progreso académico, que los estudiantes hacen de año en año en un distrito, escuela o maestro.

Dichos modelos pretenden rastrear el crecimiento anual en el aprendizaje de los estudiantes. Sin embargo, este crecimiento se modela como una serie de ganancias en el rendimiento de cada estudiante, teniendo en cuenta cada uno de los profesores que les dio instrucción, y las escuelas que proporcionaron el contexto en su aprendizaje. Así, el modelo intenta atribuir el cambio en el rendimiento de los estudiantes, a los proveedores específicos de la instrucción durante un período de tiempo específico. Este sistema de Evaluación de Valor Añadido se enfocó en el progreso de los estudiantes a través del tiempo, en lugar de su nivel de logro actual.

[5]Más información sobre el modelo EVVAS en su página web: http://www.sas.com/govedu/edu/k12/evaas/index.html

Todo esto hace que el EVAAS sea considerado uno de los modelos más complejos de valor añadido, ya que intenta estudiar el crecimiento individual de cada alumno en diferentes ocasiones de medida, mediante el empleo de un diseño que se complica más según se van incrementando el número de mediciones, y los grupos de estudio. Dicho modelo, está diseñado para realizar un seguimiento de los alumnos cuando cambian de profesor incluso, es capaz de detectar la influencia sobre el nivel de logro de distintos docentes en un mismo curso académico (Sanders et al., 1997).

Este modelo longitudinal recopila datos de los estudiantes en las materias de Matemáticas, Ciencias, Estudios Sociales y Destrezas Lectoras y del Lenguaje, desde el tercer al octavo grado. Pero durante el año escolar 2015-16, la mayoría de los distritos suspendieron la realización de dichas pruebas, ya que consideraban que no eran factibles debido a todo el proceso que suponía. Aun así, los distritos tenían la opción de continuar con las pruebas de los grados de 3º a 8º solamente, si contaban con todos los materiales para un determinado grado y asignatura. Como resultado de dicha situación, para el curso 2015-2016, las puntuaciones de escala no estuvieron disponibles, de tal manera que no se proporcionaron datos del EVAAS en las calificaciones de las distintas asignaturas. En la evaluación 2016-2017 se planteó una serie de cambios, por ejemplo, una evaluación estatal desde 3º a 8º grados en formato papel y lápiz con opción online, para los exámenes de fin de curso en escuelas de Educación Secundaria. Además, las evaluaciones incluyen preguntas más rigurosas que miden la escritura de los estudiantes, el pensamiento crítico y las habilidades para resolver problemas. Siendo todas las evaluaciones administradas a final de año o al final del semestre, en el caso de los estudiantes de la escuela secundaria.

La variable de entrada en este modelo es una matriz de datos de los alumnos en los diferentes cursos, asignaturas, profesores y centros educativos. Esto hace que se traten de modelos muy complejos, donde se estima una puntuación en cada ocasión de medida, y la varianza asociada a las medidas entre alumnos, profesores y centros de forma cruzada. El cambio se estima como la diferencia entre las mediciones teniendo en cuenta las varianzas diferentes (Navarro, 2013).

TVAAS representa un cambio de paradigma para los educadores y formuladores de políticas, ya que, al identificar las prácticas más eficaces y las prácticas menos efectivas, permite a los educadores recibir retroalimentación personalizada, la cual podría aprovecharse para mejorar las experiencias académicas de sus estudiantes.

Este modelo a pesar de su complejidad ha creado un gran interés por diversos motivos:

- No incluye variables contextuales, ya que se considera que el sujeto ejerce su propio control al utilizarse distintas medidas de logro (Sanders & Wright, 2008).
- Se encuentra diseñado específicamente para la mejora escolar mediante el estudio del crecimiento.
- Permite el uso de datos perdidos gracias a la metodología utilizada.
- Permite evaluar programas, centros, distritos, etc.
- Utiliza un tipo de estimaciones denominadas estimaciones BLUP, que se estudiaran más adelante, que son eficaces para analizar el efecto de los docentes.
- Extrae la puntuación de rendimiento previo de la puntuación de logro actual.

En definitiva, los MVA aportan una información muy relevante para la mejora escolar, pero es importante el uso que los responsables del proceso de enseñanza-aprendizaje hacen de la misma. Sobre todo, si se usa con fines de sanciones o de incentivos, ya que dicha información no es lo suficientemente precisa y requiere de otras herramientas educativas, que proporcionen un carácter más humanista a los resultados, por ejemplo, las revisiones y los controles realizados por el director del centro con ayuda de los docentes, que son los que mejor conocen lo que ocurre en sus aulas. La combinación de ambos aspectos puede proporcionar grandes mejoras en los sistemas educativos.

B) Toma de decisiones en los procesos de rendición de cuentas en educación (Accountability)

Uno de los mayores cambios en la reforma educativa durante la última década, ha sido el paso de los sistemas de rendición de cuentas basados en el cumplimiento, donde se esperaba que los educadores acataran con las reglas y regulaciones gubernamentales, hacia sistemas de rendición de cuentas basados en resultados, en los cuales los maestros son los responsables del aprendizaje de los estudiantes ante el público en general (Anderson, 2005).

El cambio hacia los sistemas de rendición de cuentas basados en los resultados, se fundamenta en la idea de que la medición objetiva del desempeño estudiantil, es la mejor manera de medir el trabajo de los maestros y las escuelas, y de motivar a los estudiantes hacia un mejor rendimiento.

Existe un tremendo interés en desarrollar modelos estadísticos de crecimiento, que usen las calificaciones de los estudiantes en las pruebas estandarizadas, para estimar los efectos que los maestros tienen sobre el aprendizaje de los estudiantes. En este contexto, los sistemas de rendición de cuentas pueden beneficiarse de los MVA, ya que dichos modelos permiten la obtención de resultados más ajustados a las características de las escuelas y de las familias.

La rendición de cuentas basada en los resultados, puede desglosarse en dos subcategorías: sistemas de rendición de cuentas fundamentados en el logro (modelos de estatus) y los asentados en el crecimiento (modelos de valor añadido). La creciente popularidad de MVA se debe en parte a la imagen más clara del aprendizaje que representan estos modelos, ya que estudian el progreso académico de los estudiantes en distintas ocasiones de medida, en vez del logro de un mismo estudiante medido en un único momento temporal (modelos de estatus).

Una crítica válida para el enfoque de rendición de cuentas basado en el logro, es que compara cosas que no son necesariamente comparables, debido a que las poblaciones de estudiantes pueden ser significativamente diferentes de un año a otro (Braun, 2005). En este sentido, los MVA al basarse en distintas mediciones que se realizan sobre el mismo grupo de sujetos, proporcionan medidas más precisas y

fiables del nivel de logro de los centros, pudiendo informar de hasta qué punto las escuelas mejoran el nivel de rendimiento de sus alumnos.

Una segunda aportación que los modelos de valor añadido pueden hacer a los sistemas de rendición de cuentas, frente a los modelos de estatus, es que los modelos de logro no tienen en cuenta que los estudiantes comienzan cada año escolar con diferentes niveles de rendimiento.

Evaluar el desempeño de un docente al examinar los niveles de logro académico de los estudiantes puede ser engañoso, puesto que algunas aulas pueden estar formadas por estudiantes que entran con altos niveles de logro, mientras que otras, pueden estar compuestas por estudiantes con una tasa de rendimiento inicial muy baja a la del resto de compañeros.

Los sistemas de rendición de cuentas basados en logros, no pueden explicar situaciones como que una clase de estudiantes con bajos niveles de rendimiento, obtengan grandes ganancias, sin alcanzar la puntuación final requerida. Tampoco pueden explicar cómo una clase con estudiantes que entran con altos niveles de logro y hacen pocos progresos, se mantienen por encima de la puntuación de corte en la competencia requerida al final del año. Tal y como se ilustra en estos escenarios, los sistemas de rendición de cuentas basados en logros, pueden castigar erróneamente a los maestros responsables de los grandes resultados educativos y premiar a los responsables de los malos resultados educativos (Linn, 2005).

En definitiva, debido a que los sistemas de rendición de cuentas basados en logros comparan diferentes poblaciones, y no controlan el rendimiento previo, son una medida pobre del efecto de la contribución de un maestro individual al logro estudiantil. Por el contrario, los MVA están diseñados para evaluar el crecimiento del aprendizaje en una población de estudiantes, supervisar factores más allá del control del profesor, y aislar las contribuciones que los profesores hacen en las ganancias de los estudiantes. Como resultado, los MVA proporcionan una medida que permite comparaciones entre diferentes profesores, y se pueden extraer inferencias tales como que el profesor X estimula más aprendizaje que el profesor Y. Las comparaciones permitidas por los MVA proporcionan una medida más justa y objetiva del desempeño del maestro que aquellos sistemas que se basan en el logro.

Los partidarios de los MVA argumentan que estos proporcionan un medio justo y útil a la hora de diferenciar el desempeño entre los maestros, algo que los sistemas escolares tradicionalmente han sido reacios o incapaces de hacer (Goldhaber, 2010).

Una consecuencia de las evaluaciones basadas en el logro, es que los profesores se pueden desmoralizar si perciben que el sistema de evaluación es injusto o arbitrario. Los datos de algunas encuestas indican que las presiones de rendición de cuentas están asociadas con un mayor desgaste y una moral reducida, especialmente entre los docentes de las escuelas desfavorables (Feng, Figlio & Sass, 2010). Paradójicamente, esto supone que, en lugar de motivar un mejor desempeño de los maestros, es posible que estos resultados injustos hagan que profesores con talento dejen la profesión por completo (Baker et al., 2010). Pues como bien lo expresó Braun (2005) un sistema de rendición de cuentas que injustamente penaliza a las escuelas que están contribuyendo al desarrollo estudiantil, incita al abandono de los maestros.

Aunque los MVA son una gran mejora respecto a los modelos de logro debido a que posibilitan un sistema de rendición de cuentas más contextualizado, y, por lo tanto, más justos., sus resultados deben utilizarse con fines informativos de cara a la mejora de los centros, ya que, si se usan como medio para el establecimiento de sanciones severas o incentivos, se distorsiona la finalidad de dichos modelos, y las escuelas y los docentes intentarán a toda costa evitar los malos resultados de las pruebas. Por ejemplo, si se recompensa el incremento del rendimiento de los estudiantes, los centros escolares podrían preparar a sus alumnos para la superación de ciertas pruebas dejando de lado importantes contenidos educativos (Baker et al., 2010).

El uso de este tipo de información proporcionada por los MVA, depende de la política educativa subyacente que se encuentre presente en el sistema educativo que se desea evaluar. En cualquier caso, hay que tener presente que dicha información procede de análisis estadísticos que pueden estar afectados por diversos problemas metodológicos (Baker et al., 2010), lo que conlleva a la necesidad de ser muy cautelosos a la hora de interpretar los mismos y a la recomendación de no utilizarse como única herramienta para la toma de decisiones sobre un centro, especialmente

si estas son de alto impacto, ya que se asumiría una relación causal entre los resultados obtenidos en los MVA y lo que realmente ocurre en las escuelas.

C) Elección de centros escolares

Vinculado con los sistemas de rendición de cuentas, se encuentra el uso de las publicaciones de los resultados académicos, obtenidos a partir de la estimación de los MVA, con la finalidad de facilitar a las familias la elección de centros educativos para sus hijos. Esto conlleva a la publicación del rendimiento que tienen las diferentes escuelas en las evaluaciones realizadas, como medio de ayuda en la elección de escuelas.

Aunque en muchos países las familias no pueden elegir el centro al que van a estudiar sus hijos, ya que son destinados directamente a la escuela local o bien no es posible por las altas exigencias de los centros. En el caso de que puedan ejercer su libre derecho en la elección de la misma, la metodología del VA proporciona a las familias información enriquecedora acerca del estado académico de las diferentes escuelas.

El comienzo de la publicación del nivel de logro de los centros educativos se inicia con la normativa *Every Child Matters* (1988)[6], donde la finalidad de la evaluación era la identificación de centros eficaces e ineficaces, para la introducción de mejoras, y facilitar a las familias la elección del centro educativo. Esta información se proporcionaba en las controvertidas *Tablas de Liga*, que fueron altamente criticadas, ya que se consideraban que sus resultados brutos eran muy injustos, debido a que se encontraban muy afectados por el rendimiento previo de los estudiantes, y por variables sociodemográficas (Saunders, 1999; Yang, Goldstein, Rath & Hill, 1999). Así, Goldstein y Spiegalhalter (1996) demostraron que, si se ajustaba dichas puntuaciones, teniendo en cuenta los niveles de logro previo de los estudiantes, cambiaba la clasificación de las escuelas. Además, cuando se utilizaba los niveles de confianza respecto a la media, la mayoría de los centros educativos tenían promedios muy similares (Fitz-Gibbon, 1997). Esto dio lugar al desarrollo de aproximaciones evaluativas basadas en MVA, con la intención de ajustar el

[6] Para más información consultar Reid (2005)

rendimiento anterior y posterior al VA contextualizado, incluyendo intervalos de confianza (Ray, Evans & McCormack, 2009).

Si los sistemas de rendición de cuentas utilizan los resultados de sus publicaciones con fines informativos, y de ayuda a las familias en la elección de los centros educativos, los datos que proporcionan los modelos de VA son mucho más justo que los resultados brutos. Aun así, hay que ser cautelosos en la difusión de la información, sobre todo cuando se habla de ordenaciones o ranking de centros, ya que, generalmente, existen escasas diferencias significativas entre las escuelas, debido fundamentalmente al error muestral. Además, el uso de MVA suponen la utilización de distintas medidas en el tiempo que se encuentran afectadas por el error de medición haciendo que la correlación entre las mismas no sea perfecta, lo que origina la regresión o egresión a la media, que puede afectar a la ordenación de los centros, si no se trabaja con un número de medidas suficientes.

Tras todo lo comentado anteriormente en dicho apartado, el VA supone una herramienta muy útil y beneficiosa en el campo educativo, que puede ayudar a la toma de decisiones y a introducir importantes mejoras. Todo ello depende del uso y cómo se interprete la información que dichos modelos proporciona, pero aun así ay que tener en cuenta que los mismos no son perfectos y no carecen de dificultades y/o problemas.

I.5.4. Problemas y dificultades del valor añadido

Martínez-Arias (2009) plantea que los problemas vinculados con los MVA son de carácter fundamentalmente estadístico, psicométrico y práctico. Entre los problemas estadísticos se puede destacar los **supuestos del modelo**. Los métodos de estimación de los MVA requieren de procedimientos paramétricos que exigen que se den una serie de supuestos. Generalmente, no suele existir un modelo que cumpla con todos ellos y, por lo tanto, que sea perfecto. Lo importante es determinar los sesgos que introducen el incumplimiento de alguno de los supuestos, ya que la violación de los mismos puede afectar de distinta manera a los resultados obtenidos (Martínez-Arias, 2009).

En los MVA es muy común el uso de covariables, generalmente puntuaciones de los test, como predictores del modelo, en este caso es importante tener en cuenta el supuesto de que dichas covariables deben estar medidas sin error. El hecho de trabajar con puntuaciones de tests que presentan errores de medida, puede afectar a la estimación de los efectos de la escuela. En esta situación es conveniente el uso de las puntuaciones realizadas durante varios años, ya que pueden reducir los efectos (Lockwood & McCaffrey, 2009).

Los **datos perdidos** también se relacionan con los problemas estadísticos de estos modelos. Generalmente, los MVA requieren de muchos datos por parte de los estudiantes, puesto que usan las puntuaciones de un mismo sujeto en diversos momentos temporales. Esto da lugar a la existencia habitual de registros incompletos que en los modelos univariantes supone la eliminación de aquellos casos con datos faltantes. Sin embargo, los modelos multivariantes cuentan con la ventaja de que no requieren de una matriz de puntuaciones completa, y, por lo tanto, se puede trabajar con una base de datos que cuenten con datos perdidos, siempre y cuando, tengan una distribución aleatoria (Martineau, 2009). También, se pueden aplicar técnicas de imputación de datos perdidos, pero, independientemente de la aproximación utilizada, la existencia de datos perdidos puede conllevar a importantes sesgos a la hora de estimar los efectos de las escuelas.

El **tamaño muestral** también es otro problema estadístico a considerar. Una muestra pequeña introduce una mayor incertidumbre en la estimación del valor añadido de las escuelas, debido a la estabilidad reducida de las puntuaciones (OECD, 2011b) esto provoca medidas menos fiables y precisas. Generalmente, las evaluaciones realizadas con cohortes con menos de 20 alumnos producen estimaciones de VA que dificultan la interpretación de los resultados. Además, cuando se utilizan modelos multinivel, tal y como apuntan diversos autores (Bryk & Raudenbush, 1987; Goldstein, 1997; Raudenbush & Bryk, 2002; Snijders & Bosker 2012), se recomienda que el número de estudiantes por aula o centro sea superior a 20 para que no se produzcan sesgos a la hora de estimar la pendiente media de crecimiento, aunque no se han encontrado investigaciones que demuestren cómo afecta el tamaño muestral en el error de estimación de esta pendiente de crecimiento.

La estabilidad de las puntuaciones, así como la elección del MVA son aspectos que también pueden afectar a las estimaciones de las escuelas, ya que diferentes modelos darán lugar a diferentes estimaciones del error, que acompañan a las puntuaciones de VA de los centros educativos (OECD, 2011b). Estos errores también se pueden ver afectados por la inclusión de variables socioeconómicas que muchas veces no se incluyen en los modelos.

En cuanto a los problemas psicométricos, la mayoría de ellos están relacionados con el uso de **escalas verticales**. Para poder comparar los resultados a lo largo del tiempo es necesario disponer de medidas escaladas verticalmente, tal y como se explicará en el siguiente capítulo.

La creación de escalas verticales requiere de procedimientos de equiparación de las distintas formas del test, con la finalidad de comparar las puntuaciones.

Existen dos formas de equiparación. La equiparación horizontal que se emplea cuando se cuenta con formas múltiples, que tienen dificultad similar y evalúan contenidos parecidos dentro del mismo curso o edad (Martínez-Arias, 2009), y la equiparación vertical que se utiliza para relacionar las puntuaciones de los test que tienen distinta dificultad (generalmente por ser de distintos cursos), pero evalúan los mismos contenidos en sentido general. El escalamiento vertical en los MVA es de gran importancia porque las inferencias realizadas del VA depende de la adecuación de la escala, que generalmente va empeorando según aumenta la distancia entre los puntos temporales.

Respecto a los problemas prácticos, están más vinculados con la **administración de los test.** En este sentido existen dos cuestiones claves como son el número de medidas necesarias y las fechas de recogida de la información. Respecto al número de medidas se requiere un mínimo de dos mediciones, generalmente dos cursos consecutivos, para estudiar el VA. En estos modelos, se analiza la diferencia entre ambos cursos como ganancia entre las dos ocasiones de medida. Aunque estos modelos de ganancia, tal y como indicó Rogosa (1995), son mejores que los que tienen una sola medida no los mejoran mucho. Lo ideal es el uso de modelos con mediciones múltiples, ya que permiten un tratamiento multinivel y multivariante, donde se puede estudiar el crecimiento de los estudiantes y las escuelas de manera más precisa (Raudenbush, 2004, Raudenbush & Bryk, 2002).

En cuanto al momento de recogida de información, es un tema importante, pues el intervalo exitente entre las tomas de datos suele ser bastante amplio, incluso a veces interumpido por el verano, lo que hace que su efecto pueda vincularse con el estatus socioeconomico del estudiante (McCaffrey et al., 2003a) pudiendo afectar a la validez de las estimaciones.

Otro aspecto de interés previo a la aplicacion de los test, esta vinculado con las cuestiones de planificación de la evaluación.

Durante la planificación de la misma es necesario la colaboración de toda la administracion educativa, todos los agentes implicados deben comprometerse en la participacion activa en la misma. También, es importante desde el principio tener claro cuales son los objetivos y el uso que se va hacer de los resultados obtenidos. En cuanto a los costes, cuestión importante en el mundo educativo, este dependerá fundamentalmente del tamaño de la muestra, del número de mediciones, formato de recogida de la información, etc.

Durante el diseño de las evaluaciones es necesario tomar decisiones respecto al contenido a evaluar y la forma de hacerlo, por ejemplo, si se usan ítem de respuesta construida o ítems cerrados. Es muy importante la identificación de los alumnos de tal manera que se puedan seguir a lo largo del tiempo. Además, al utilizar distintas puntuaciones de logro, es necesario definir cómo se van a comparar esas puntuaciones en las distintas ocasiones de medida, algo que influye en el diseño de los propios instrumentos de medida.

Por último, a la hora de presentar los resultados es importante tener en cuenta que deben dar respuesta a los objetivos planteados previamente y se debe realizar una interpretación adecuada y clara de los mismos, ya que el uso de modelos complejos puede hacer que la información sea menos entendible para aquellas personas que no tengan altos conocimiento sobre estadística. Además, la transparencia es el indicador que debe guiar toda la evaluación.

Todos estos problemas no deben pasar desapercibidos, debido a la compleja dificultad y repercusiones que supone clasificar a las escuelas en términos de alto y bajo desempeño. Muchos de ellos se estudian con mayor profundidad en el siguiente capítulo (capítulo II) que trata sobre los principales aspectos metodológicos de los modelos de valor añadido.

En definitiva, los modelos de valor añadido basados en los modelos de crecimiento, son una alternativa mucho más eficiente y eficaz que otro tipo de modelos como son los de cambio de cohorte o de ganancia, ya que aportan mayor información de carácter más preciso. Pero aun así, no deben ser considerados como la panacea, puesto que no se encuentran exentos de problemas metodológicos que pueden afectar a la estimación de los resultados, por lo que, no se deben utilizar como una única herramienta en la toma de decisiones para la mejora del sistema escolar, sino que deben combinarse con otra información proporcionada por los agentes educativos que trabajan y conocen a los estudiantes.

CAPÍTULO II. ASPECTOS METODOLÓGICOS DE LOS MODELOS DE VALOR AÑADIDO

Aunque la mayoría de los estados americanos, mejoraron sus sistemas de recogida y procesamiento de datos en respuesta a la ley *No Child left Behind Act* (McCall et al., 2004), todavía en muchos casos, las exigencias y requisitos metodológicos necesarios para la consecución de estimaciones fiables de VA, han sido obviadas.

En la terminología estadística, a lo primero que se debe prestar atención es a la estructura de la información disponible, ya que los estudiantes están anidados dentro de las escuelas y a su vez las escuelas se pueden encontrar anidadas dentro de distritos o comunidades. Esta estructura jerárquica conlleva el uso de métodos multinivel para un tratamiento adecuado de los datos (ver Aitken & Longford, 1986; Brown, 1994; Burstein, 1980; Cronbach & Webb, 1975; Goldstein, 1988; Raudenbush & Willms, 1995).

Por desgracia, el estudio del cambio no es tan sencillo como a muchos investigadores les gustaría, ya que, además de métodos específicos para el análisis de estructuras anidadas, el estudio del cambio a través del tiempo supone el empleo de procedimientos complejos para la comparación de puntuaciones.

Históricamente, en las ciencias sociales, se ha intentado estudiar el cambio mediante diseños longitudinales basados en observaciones realizadas a un individuo en un pre-test y post-test, de tal manera que el crecimiento solo podía ser evaluado de manera lineal como ganancia entre las aplicaciones.

Pero la evidencia de que el rendimiento escolar no puede estimarse adecuadamente cuando los resultados de las pruebas del estudiante se agrupan en uno o dos puntos temporales, o cuando sucesivas cohortes de estudiantes se comparan, ha llevado a varios autores a argumentar a favor del uso de análisis longitudinales, basados en al menos tres mediciones realizadas a un mismos sujeto a lo largo del tiempo (Barton & Coley, 1998; Bryk & Raudenbush, 1988; Linn & Haug, 2002), de tal manera que tanto el desempeño de los estudiantes, así como la estimación de los efectos escolares, se puedan estimar de manera más precisa y directa.

Al utilizar las puntuaciones de sujetos evaluados en distintos momentos temporales, se requieren una serie de procedimientos que permitan la comparación de los diferentes resultados obtenidos a lo largo del tiempo. Para ello la construcción de escalas comunes, mediante métodos de equiparación y/o escalamiento, que faciliten la comparación de puntuaciones es una exigencia fundamental de los MVA. Generalmente, el método utilizado en los MVA para la comparación de puntuaciones en las distintas ocasiones, es el escalamiento vertical, ya que permite comparar los resultados de pruebas realizadas en distintos grados académicos.

Estos modelos de crecimiento con carácter anidado y basados en modelos longitudinales, pueden resumir de manera más realista y eficiente la información disponible en las evaluaciones educativas. Debido a que identifican la trayectoria de crecimiento individual de cada estudiante y permiten un cierto control sobre problemas métricos y factores de confusión, ya que según Stevens y Zvoch (2006) cuentan con la ventaja de que cada estudiante desempeña su propio control. De tal manera que la curvatura general de la trayectoria de crecimiento de un estudiante, se convierte en un índice útil del cambio en los aprendizajes, lo que conduce en determinadas ocasiones a hablar de causalidad entre las puntuaciones estimadas y la calidad del aprendizaje.

Sin embargo, autores como Raudenbush y Bryk (2002) consideran que en la estimación de los MVA es fundamental la introducción de ciertos predictores como el nivel socioeconómico, el rendimiento previo, el sexo, etc., ya que ayudan a controlar el efecto que estas variables tienen sobre las estimaciones finales. Aun así, la introducción de dichas covariables, no permite establecer relaciones causales, ya que, a diferencia de los diseños experimentales puros, en estos modelos no se realiza una asignación aleatoria de los estudiantes a los centros, lo que hace que muchas variables no se controlen, y, por lo tanto, sea un tanto osado hablar de causalidad, pues la selección no aleatoria de los estudiantes hace que las escuelas cuenten con características de los mismos distribuidas de manera desigual. Esto supone contar con centros con un gran porcentaje de alumnos desventajados y otros con un gran número de alumnos aventajados, lo que puede suponer un problema en la rendición de cuentas, ya que las escuelas con un alto porcentaje de alumnos desfavorecidos

corren el riesgo de sufrir evaluaciones desfavorables, sino se tiene en cuenta ciertos aspectos metodológicos.

La elaboración y estimaciones del VA, supone el cumplimiento de una serie de cuestiones metodológicas, que si no son tratadas de manera correcta harán que los análisis de VA valoren de manera incorrecta la eficacia de muchos profesores y escuelas, lo que podría dificultad los esfuerzos realizados para mejorar los sistemas educativos (McCaffrey, Lockwood, Koretz & Hamiltion, 2003b). No obstante, todos los análisis del VA cuentan con aspectos que pueden amenazar la validez de las estimaciones de los efectos de las escuelas o de los profesores, por lo que, hay que ser cautelosos con el uso que se hace de los resultados obtenidos mediante dichos modelos, debiendo utilizar los mismos con fines más descriptivos que causales.

Desde esta perspectiva, en este segundo capítulo se pretende realizar un recorrido por los aspectos metodológicos más importantes de los MVA, para ello se estudiará la estructura anidada de los datos, su carácter longitudinal, los procedimientos utilizados para la comparación de puntuaciones a lo largo del tiempo, y la importancia de las variables contextuales en este tipo de modelos.

II.1. Anidamiento de los datos

Previo al estudio de los dos principales modelos estadísticos existentes para el análisis del valor añadido: modelos jerárquicos lineales y modelos lineales mixtos (capítulo III), es necesario describir por qué dichas técnicas son las más adecuadas en el análisis de los datos procedentes de las evaluaciones educativas a gran escala.

Históricamente, la investigación educativa ha tendido a evaluar los efectos de los estudiantes, profesores, aulas y otras variables educativas, utilizando técnicas de regresión de un solo nivel. Se ha utilizado un enfoque transversal que ignoraba la naturaleza anidada de los datos educativos (Lee, 2000), lo que daba lugar a conclusiones erróneas al considerar las observaciones independientes cuando realmente eran dependientes. Aumentando, de tal manera, la probabilidad de cometer un error de Tipo I (Singer & Willett, 2003).

Además, dichas técnicas transversales asignan el mismo valor de las variables macro, como puede ser el aula o la escuela, a las unidades micro, por ejemplo, el estudiante, dejando de lado la posible variación de dichos factores entre

sujetos y dando lugar a lo que bien denominó Alker en 1969 como falacia atomística. Del mismo modo, también es frecuente en estas técnicas la presencia de la falacia ecológica (Robinson, 1950) al otorgar el valor medio de una variable del nivel de sujeto al nivel de grupo. Esto es posible en el estudio de relaciones a nivel macro, pero extrapolar las conclusiones al nivel del individuo es un error.

Así, el uso de herramientas tradicionales como ANOVA y mínimos cuadrados ordinarios (MCO), pueden producir resultados espurios, ya que generalmente, en el campo educativo los estudiantes que van a la misma escuela son más parecidos que los estudiantes que pertenecen a diferentes centros educativos (Tabachnick & Fidell, 2007). En este contexto, la estructura de datos es muy probable que contenga tres o más niveles (por ejemplo, medidas repetidas anidadas dentro de alumnos y estudiantes anidados dentro de las escuelas). Sin embargo, a pesar de estas consideraciones algunos investigadores han analizado y siguen analizando sus datos educativos mediante modelos generalizados, sin tener en cuenta la estructura jerárquica del nivel superior (por ejemplo, escuelas) y asumiendo la independencia entre los estudiantes (por ejemplo, Boscardin, Muthén, Francis & Baker, 2008; D'Angiulli, Siegel & Maggi, 2004).

El grado de homogeneidad de los contextos está definido por la autocorrelación o correlación intraclase, que si no se tiene presente puede dar lugar a las siguientes consecuencias (Gaviria & Castro, 2005):

a) Como los estudiantes que van a la misma escuela tienden a parecerse entre ellos, la información a nivel individual es menor que la que pueden proporcionar los estudiantes que proceden de distintos centros.
b) Como los test se basan en el supuesto de independencia de las observaciones, los errores típicos son muy pequeños. Aunque en dicha estructura anidada este supuesto no se cumple, y, por lo tanto, es probable que se lleguen a resultados significativos cuando realmente no lo son.

Cuando no se utiliza algún tipo de modelado multinivel con datos jerárquicos, estos hechos se ignoran, y pueden incapacitar la validez de los resultados de un estudio (Bickel, 2007).

Un ejemplo práctico de los posibles sesgos que pueden ocurrir cuando no se utilizan las técnicas de análisis adecuadas a las características de los datos, aparece ejemplificado en un estudio llevado acabo por Richter (2006) quien compara distintos métodos para el análisis de los datos jerárquicos referido al tiempo de lectura. Para ello Richter utiliza cuatro técnicas estadísticas: (a) un modelado jerárquico lineal con dos niveles (nivel 1: oraciones y nivel 2: personas), (b) una ANOVA de factores mixtos basado en los tiempos de lectura agregados entre frases, (c) un análisis de regresión múltiple donde los tiempos de lectura de oraciones son tratados como observaciones independientes, y (d) un análisis de regresión múltiple basado en modelos de regresión distintos para cada persona.

Los resultados de la ANOVA arrojaron sólo un efecto principal (frases plausibles), mientras que los modelos jerárquicos lineales mostraron dos (frases plausibles y objetivo de la lectura). En consecuencia, dos de los efectos que fueron significativos en el análisis de los modelos jerárquicos lineales, no se detectaron por el ANOVA. Esto se debió a la pérdida de información producida por usar un único nivel, y al fracaso del ANOVA para separar la varianza correspondiente al nivel de persona y el nivel de oración de una manera apropiada.

Respecto a los análisis de regresión múltiple donde los tiempos de lectura de las oraciones eran tratados como observaciones independientes, algunos resultados fueron sobreestimados, por ejemplo, la pendiente para el efecto principal de la plausibilidad de la oración y otros subestimados, como la interacción entre la plausibilidad de la oración y la meta de la lectura. Los errores estándar de diversos parámetros, el intercepto y los efectos principales para el objetivo de la lectura y el número de sílabas, así como la interacción entre la meta de lectura y el número de sílabas, fueron también subestimados, llegando a producir un falso error positivo. Estas diferencias respecto a los resultados obtenidos en el modelo jerárquico lineal, se deben a que las estructuras de los errores de los análisis de regresión múltiple, son menos complejas y como consecuencia los dos efectos aleatorios que fueron significativos en el modelo jerárquico lineal, no pueden ser incluidos en un análisis de regresión múltiple de un solo nivel, ya que se produce un modelo mal especificado.

En el caso de los análisis de regresión múltiples realizados por separado para cada participante, no detectaron la significatividad de dos de los efectos, y sobreestimaron ligeramente los resultados respecto al modelo jerárquico lineal. Del mismo modo, los errores estándar también se sobreestimaron significativamente, lo que pudo haber conducido a falsos errores negativos en las pruebas de contraste de hipótesis. Aun así, de todos los modelos analizados, dicha estrategia metodológica fue la que se consideró más afín al modelo jerárquico lineal.

A través de este ejemplo, desde una perspectiva tanto metodológica como epistemológica, se muestra como la falsa conclusión de que una hipótesis no es válida, puede ser tan perjudicial como aceptar falsamente una hipótesis como válida. Por todo esto, es conveniente utilizar métodos de análisis adecuados a las características de los datos disponibles, que permitan garantizar la validez de los resultados, en el caso de datos anidados los métodos multinivel son la opción más correcta.

Recientemente, en la última década los investigadores del campo de la educación han comenzado a utilizar estas técnicas que tienen en cuenta varios niveles (técnicas multinivel), como procedimientos de referencia para el análisis de datos con estructuras jerárquicas complejas (O'Connell & McCoach, 2008).

El uso de dichas técnicas se encuentra totalmente justificado, ya que la disciplina educativa se caracteriza por una relación de pertenencia anidada entre las unidades de dicho sistema. Pues, tal y como se ha comentado en este mismo apartado, se puede encontrar a estudiantes agrupados en aulas, dichas aulas pueden agregarse dentro de las organizaciones educativas, y estas organizaciones pueden anidarse en distritos, comunidades o países. Esto supone que no se puede estudiar un determinado nivel sino se tiene en cuenta el resto, por lo tanto, la recolección de datos de las evaluaciones correspondientes debe reconocer y recoger el carácter anidado propio de los sistemas educativos; para ello los investigadores han comenzado a utilizar procedimientos de investigación que tienen en cuenta los distintos niveles de anidación.

Estas técnicas utilizadas con datos de carácter anidado, se han etiquetado como modelos multinivel lineal, modelo jerárquico lineal, modelos mixtos, modelos agrupados, y/o modelos de coeficientes aleatorios, entre otras muchas. Lo que cada

uno de estos términos tiene en común, es que permiten que los coeficientes de regresión (intersecciones y pendientes) puedan variar a través de las unidades de anidación como pueden ser el tiempo, el estudiante y el aula (Raudenbush & Bryk, 2002). Además, en dichos modelos aparecen identificados las relaciones que se producen tanto dentro de un mismo nivel como entre niveles, ya que se trabaja con todos los niveles a la vez, siendo posible diferenciar la varianza que cada predictor explica en los distintos niveles jerárquicos[7].

II.2. Datos longitudinales

En los MVA, la inclusión de distintas medidas de logro es su marca de clase. Tanto es así que los modelos de valor añadido, no se pueden entender fuera de un diseño longitudinal, y sin un referente temporal, dentro de la definición del modelo más básico (Ruiz de Miguel & Castro, 2006).

Las evaluaciones basadas en un único punto temporal son consideradas a menudo deficientes, si lo que se pretende es estudiar las diferencias en los niveles de logro de las escuelas, ya que se pueden confundir los procesos de instrucción y de la calidad de enseñanza en los centros educativos (Tekwe et al., 2004). Por ello, con el fin de conocer con mayor exactitud el crecimiento de las escuelas y las diferencias existentes entre las mismas, es más adecuado el uso de distintas puntuaciones temporales que permitan realizar un seguimiento del cambio que se produce en el desempeño académico de los estudiantes a lo largo del tiempo.

Para medir el crecimiento individual del estudiante, se necesita establecer un sistema de datos longitudinal, puesto que permite seguir a los mismos estudiantes a medida que avanzan a lo largo de los cursos académicos. También, posibilita seguir a los estudiantes de escuela en escuela e incluso de un distrito a otro dentro de un estado.

El ingrediente clave de un sistema de datos longitudinal en los modelos de valor añadido, es el uso de un identificador único para cada estudiante, o más comúnmente, un número de identificación del estudiante. Un identificador de estudiante funciona exactamente igual que un número de la seguridad social. A cada

[7] Estos modelos serán estudiados con mayor detalle en el Capítulo III: Estimación del Valor Añadido en las Evaluaciones Educativas.

estudiante se le asigna un número único cuando entran en el sistema escolar y permanece con el mismo número durante toda su carrera académica, incluso se puede mantener cuando cambian de escuela o se mudan a un nuevo distrito en el estado. Aunque la asignación de un identificador puede parecer algo muy lógico y sencillo, la propia experiencia metodológica demuestra que en numerosas ocasiones no lo es tanto, y, a veces, sobretodo en investigadores nóveles, se produce este tipo de descuido en el cual no se identifican a los sujetos a lo largo del tiempo, lo que puede conllevar la invalidación de los resultados.

Junto con estos identificadores, es recomendable también la inclusión de características socio-métricas tales como el sexo, la edad y /o la situación socioeconómica de cada estudiante, ya que dicha información es importante para la rendición de cuentas basada en los subgrupos de estudiantes y para los modelos de valor añadido, tal y como se estudiará más adelante en este mismo capítulo.

Poco a poco, en la última década, el número de estudios longitudinales se han ido incrementando de manera notoria en las diversas disciplinas de las ciencias sociales (Khoo, Oeste, Wu & Kwok, 2006). Esto se debe a que dichos diseños, permiten evaluar la trayectoria del crecimiento de los alumnos durante un periodo de tiempo, proporcionando numerosas ventajas frente a otro tipo de diseños más tradicionales (por ejemplo, estudios de cohortes sucesivas). El potencial de estos diseños es tan útil, que se han llegado a considerar los más adecuados para evaluar el progreso de los alumnos y la eficacia de las escuelas (Stevens, 2005; Thum, 2003). Entre las ventajas de estos diseños, cabe destacar su adecuación a situaciones en las que se quiere controlar la influencia de las características socio-familiares de los alumnos (al ejercer el sujeto su propio control en las sucesivas ocasiones), u otros factores incontrolables, así como en aquellos estudios en los que el objetivo pase por medir un fenómeno cambiante a lo largo del tiempo, como puede ser el aprendizaje de los estudiantes. Estas características adquieren especial relevancia en los contextos escolares y en la interpretación de los índices de valor añadido, pues la distribución de alumnos en escuelas no responde a un patrón aleatorio (Raudenbush 2004; Rubin, Stuart & Zanutto, 2004)

Frente a estos diseños, se ha investigado la excesiva confianza de los sistemas de rendición de cuentas en la comparación de cohortes sucesivas de estudiantes, como una medida para el estudio del cambio. Por ejemplo, con el fin de medir y evaluar a las escuelas centrándose en el cambio que se produce de un año a otro en los niveles de logro de los estudiantes, se compara el rendimiento promedio de alumnos de sexto curso en el año 2016 con el rendimiento de los alumnos de sexto curso en el año 2017. El problema de dichas medidas es que tan sólo permite a los estados evaluar a las escuelas en términos de ganancias, en lugar de niveles de logro absolutos. Además, recientes investigaciones realizadas sobre la medición en cohortes sucesiva demuestran que las estimaciones de ganancia que se producen de un año a otro en la competencia evaluada, están afectadas en gran medida por la variación del muestreo, el error de medición, y otros factores que no están asociados con el tamaño de la escuela o la práctica escolar (Linn & Haug, 2002; Kane & Staiger, 2002). Así mismo, la falta de variación sistemática en las puntuaciones sucesivas obtenidas, pueden afectar a las evaluaciones de las escuelas realizadas por los estados, ya que los niveles de desempeño de los educandos se pueden ver influenciados por las fluctuaciones que se producen en las cohortes de los estudiantes, o por la administración de la prueba (Linn & Haug, 2002).

Por todo lo comentado anteriormente, es evidente que el rendimiento escolar no puede estimarse de manera fiable, cuando los resultados académicos de los estudiantes se basan en una sola puntuación de logro o en la medición de cohortes sucesivas.

Esto ha llevado a varios autores (Barton & Coley, 1998; Bryk & Raudenbush, 1988; Linn y Haug, 2002). a argumentar a favor del uso de análisis longitudinales del desempeño individual del estudiante, como una estimación más directa y precisa de los efectos escolares. Por ejemplo, Goldstein (1991) descubrió que la única forma de comparar escuelas se encuentra en el estudio de la cantidad de progreso, que los alumnos alcanzan durante su estancia en la escuela. El progreso del estudiante se mide comparando el cambio producido de un periodo de tiempo a otro, por ello, la metodología más adecuada para medir el crecimiento del rendimiento de los estudiantes, es a través de la estimación de las trayectorias de crecimiento

individual, mediante el empleo de modelos multinivel (Bryk & Raudenbush, 1992, 1987; Willett, 1988; Willms, 1992).

Los análisis multinivel de carácter longitudinal, basados en el desempeño individual del estudiante, permiten analizar las medidas de resultados a lo largo del tiempo y facilitan las estimaciones del beneficio adicional o "valor", que los estudiantes reciben al asistir a una escuela en particular (Boyle & Willms, 2001; Bryk & Raudenbush, 1988; Willms, 1992). De esta manera, dichos diseños facilitan las estimaciones del valor añadido al proporcionar un cierto grado de control sobre una gran cantidad de factores de confusión, que pueden intervenir en la evaluación de la efectividad de las escuelas.

Mediante el uso de modelos longitudinales, cada estudiante ejerce su propio control, frente a factores de confusión que son considerados como características del estudiante que se mantienen estables en el tiempo (Sanders & Horn, 1994; Stevens, 2000). Así factores como el estatus socioeconómico, étnia, diferencias culturales, etc., pueden controlarse en gran medida a través de la aplicación de un diseño longitudinal, donde se mida a los mismos sujetos a lo largo de varias puntuaciones temporales.

Miller (1998) define diseños longitudinales como pruebas repetidas que se realizan a lo largo del tiempo. La noción de "pruebas repetidas" no está bien definida y con frecuencia parece referirse a un mínimo de dos ocasiones, algo que no ha sido aceptado por autores como, por ejemplo, Rogosa (1988, 1995) quien considera que para hablar de diseños longitudinales es necesario al menos contar con tres ocasiones de medida, especialmente si se quiere describir el crecimiento de los estudiantes.

Concretamente, Rogosa en su obra titulada *"Myths About Longitudinal Research"* del año 1988, abordó una serie de falacias cruciales sobre el análisis del cambio. Entre ellas estableció que el error más común era considerar dos aplicaciones, como medidas suficientes para realizar un estudio longitudinal.

Es cierto que el diseño pre-test, post-test es el más común en el estudio del cambio, y dos observaciones repetidas constituyen de hecho un estudio longitudinal. Sin embargo, dos medidas no son adecuadas para el estudio de la forma del crecimiento. Dos ocasiones de medida, sólo pueden estimar la cantidad de

cambio como una línea recta que pasa entre dos puntos, pero no permite evaluar la adecuación de la línea, ni comparar la línea con otras formas funcionales de cambio.

Rogosa (1988) también demostró que la cantidad de cambio producido entre dos puntos temporales puede ser engañosa si la tasa de crecimiento no es constante, sino que depende del tiempo. En este caso, el cambio dependerá crucialmente de los tiempos de medición y de las observaciones de los individuos en las dos medidas realizadas, lo que puede dar lugar a resultados contradictorios, inadecuados para definir el crecimiento.

En resumen, aunque se puede trazar la cantidad de cambio observado entre dos puntos en el tiempo, no se puede determinar la forma de la curva de crecimiento a partir de sólo dos observaciones. Por otra parte, si la función de cambio no es una función lineal, entonces el tiempo de la medición puede ser muy influyente, por ello Rogosa (1995) recomienda el uso de múltiples puntos de medición con el fin de mejorar el análisis estadístico, y el examen de las trayectorias individuales de crecimiento a través del tiempo.

Otros investigadores (Rogosa, Brandt, & Zimowski, 1982; Willett, 1989) también han demostrado que el uso de tan sólo dos puntos temporales, es una estrategia inadecuada e insuficiente para medir el crecimiento, ya que contienen información mínima sobre el cambio individual. Además, las estimaciones realizadas utilizando tan sólo dos puntos temporales se encuentran sesgadas, ya que siempre se ven afectadas por el efecto de regresión a la media[8] (Nesselroade et al., 1980).

A pesar de todos estos inconvenientes, los primeros estudios sobre el cambio no comenzaron con análisis longitudinales que usaban más de dos tomas de datos, sino que inicialmente se utilizaban tan sólo dos mediciones, generalmente un pre-test y un post-test, con el fin de estudiar la ganancia que se producía entre ambos momentos temporales (Willet, 1994). Aunque es cierto que en las ciencias de la conducta la preocupación por los datos con tan solo dos mediciones ha sido de suma importancia, se ha reconocido suficientemente que los diseños longitudinales con solo dos aplicaciones contienen una cantidad limitada de información acerca del

[8] El Efecto de Regresión a la Media será estudiado en mayor profundidad en el capítulo 5 titulado El Efecto de Regresión a la Media en los Modelo de Valor Añadido

cambio de cada individuo, que no es suficiente para definir la trayectoria individual de los sujetos (Rogosa, 1988), y, por ende, es insuficiente para dar respuesta a importantes preguntas sobre el cambio y el crecimiento de los sujetos.

Todo esto ha originado que con el paso del tiempo el número de ocasiones de medidas se haya incrementado con el objetivo de mejorar las estimaciones y la modelización del crecimiento (Stevens & Zvoch, 2006; Zvoch & Stevens, 2003). En general, con tres o cuatro mediciones bien espaciadas, es posible mostrar la forma y dirección del crecimiento en los niveles de logro, aunque en determinados casos si la trayectoria individual del crecimiento es compleja, puede que sean necesarias un mayor número de medidas menos espaciadas (Willet, 1994).

Parece evidente que los datos con más mediciones ofrecen estimaciones más precisas que las realizadas con tan sólo dos tomas de datos, ya que según se incrementa el número de aplicaciones se obtiene mayor información adicional sobre cada individuo. Así, por ejemplo, Gray, Jesson, Goldstein, Hedger y Rabash (1995) señalan que el número de periodos necesario para identificar el verdadero crecimiento de los alumnos y de la escuela está entre tres y cinco, aunque, hasta el momento, no existe ningún desarrollo empírico que demuestre cual es el número de medidas necesarias para poder conocer el crecimiento real del estudiante, algo que se estudiará en la parte metodológica de dicha tesis.

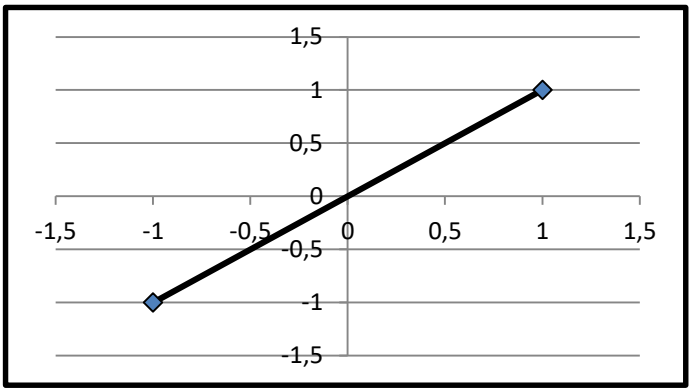
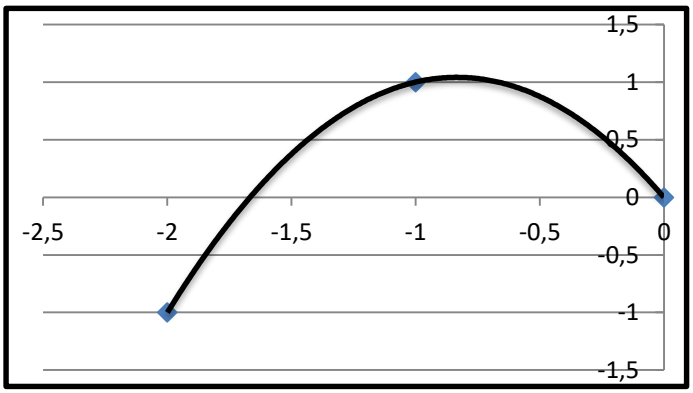
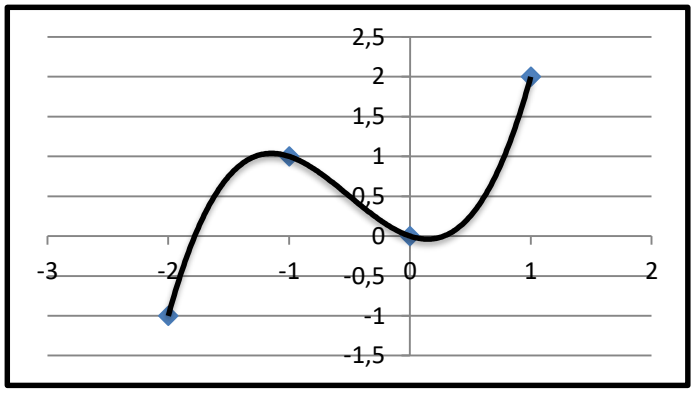
En cualquier caso, como bien apunta, Nesselroade, et al., (1980) para poder mejorar los procesos de estimación del cambio, la investigación empírica debe extenderse más allá del uso de dos puntuaciones. Pues, en los diseños en los que sólo se mide a los alumnos en dos momentos temporales no es posible evaluar la forma de la función de crecimiento, ya que sólo se puede ajustar una función lineal. Con tres oleadas de datos ya es posible evaluar el ajuste de una función lineal y de una cuadrática. Y con cuatro puntos de medida, puede ajustarse una función lineal, cuadrática o curvilínea (Stevens & Zvoch, 2006).

A pesar de esta variedad de funciones polinómicas, la mayoría de los métodos de estimación del crecimiento o cambio, presuponen que el crecimiento de cada estudiante en el logro sigue una línea recta; esto es generalmente una suposición razonable. Sin embargo, hay evidencia de que el crecimiento durante una variedad de puntos temporales puede modelarse de diferentes maneras, ya que los modelos

longitudinales permiten la evaluación del crecimiento tanto lineal como no lineal (Garson, 2013). A través de estos procedimientos, el crecimiento en el tiempo, de acuerdo al número de puntos temporales, se puede representar como un polinomio de primer grado (modelos de crecimiento lineal), un polinomio de segundo grado (modelo de crecimiento cuadrático) o como un polinomio de tercer grado (modelos de crecimiento cúbicos) o, incluso, se pueden incorporar términos de orden superior (Garson, 2013; Raudenbush & Bryk 2002).

En la siguiente Tabla II.1, se pueden observar cada una de estas funciones

Tabla II.1. *Funciones polinómicas*

FUNCIÓN	ECUACIÓN	EJEMPLO
LINEAL	$E(Y_t) = \beta_0 + \beta_1 t$ β_0 es la constante y β_1 el coeficiente de regresión asociado a la variable t (variable independiente), que en los modelos de valor añadido se refiere al tiempo.	 **Figura II.1.** *Ejemplo de función lineal* Fuente: elaboración propia
CUADRÁTICA	$E(Y_t) = \beta_0 + \beta_1 t + \beta_2 t^2$ β_0 es la constante y las pendientes de crecimientos son: β_1 y β_2, para el tiempo (t).	 **Figura II.2.** *Ejemplo de función cuadrática* Fuente: elaboración propia
CÚBICA	$E(Y_t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3$ β_0 es la constante, $\beta_1, \beta_2, \beta_3$ son los coeficientes de regresión para el tiempo, que en el caso de β_2 está elevado al cuadrado y de β_3 al cubo.	 **Figura II.3.** *Ejemplo de función cúbica* Fuente: elaboración propia

Fuente: elaboración propia

Aunque los diseños longitudinales sean el marco más apropiado para la evaluación del cambio y del crecimiento de los aprendizajes escolares, su aplicación no está exenta de problemas. Entre ellos la elección de la primera ocasión de medida

es un aspecto de vital importancia, especialmente cuando el número de mediciones es pequeño. La elección del punto de partida puede tener un gran impacto en la evaluación y en la interpretación de los resultados, ya que al ser inversamente proporcional al número de mediciones con las que se trabaje, puede afectar a la estimación del crecimiento (Stevens & Zvoch, 2006). De esta manera, la elección del punto de partida afecta menos a las estimaciones cuanto mayor sea el número de mediciones. Además, Rogosa (1995) demostró que la elección de un determinado punto de partida afecta al valor de la correlación y la intensidad, entre las distintas ocasiones de medida, lo que puede influir en el comportamiento de determinados artefactos estadísticos, como es el efecto de regresión a la media.

Con el uso de sólo dos puntuaciones de medida, no se consigue extraer todo el potencial de los estudios longitudinales, ni se permiten estimaciones precisas de la pendiente de la función de producción (Rogosa, 1979; Willet, Singer & Martín, 1998), debido a que con sólo dos observaciones las puntuaciones en el pre-test influyen en gran medida en la pendiente de la función de crecimiento (Rogosa, 1995).

La organización y gestión de los diseños longitudinales no deja de ser un asunto complejo, ya que para estimar la función de crecimiento se requiere hacer un seguimiento individual de los estudiantes a lo largo de al menos tres mediciones, lo que supone un alto coste a las políticas educativas, además de problemas relacionados con la mortalidad experimental debido, entre otras causas, a la movilidad de los estudiantes.

Otra limitación de los diseños de investigación longitudinales, es que son muy vulnerables a muchas de las amenazas a la validez comúnmente asociada con la investigación cuasiexperimental, es decir, la selección, el desgaste, la instrumentación, y todo lo que conlleva el proceso de montaje de una adecuada muestra para un estudio longitudinal, algo que no es tarea fácil (Shadish, Cook & Campbell, 2002).

Por último, algunos autores han argumentado que los diseños longitudinales solo son realmente útiles para describir el cambio, pero no para explicarlo (Baltes, Reese & Nesselroade, 1977).

A pesar de todos estos posibles inconvenientes, en este trabajo se considera que los diseños longitudinales con más de dos ocasiones de medida, suponen un gran avance metodológico, ya que permiten evaluar la trayectoria del crecimiento académico de los estudiantes, proporcionando más información y de carácter más precisa que otros diseños, como pueden ser, el de estatus, el de cambio de cohorte o incluso los modelos de ganancia. La utilización de múltiples ocasiones de medida, permite además aumentar la fiabilidad de la estimación de la función de crecimiento e incrementar la validez interna de la valoración del rendimiento (Cook & Ware, 1983; Rogosa, 1979; Rogosa, Brandt & Zimowski, 1982; Singer & Willet, 2003). Además, proporcionan un control más exhaustivo sobre determinados artefactos estadísticos (por ejemplo, efecto de regresión a la media), influencias y un mejor entendimiento de la forma de la función de crecimiento, siendo posible la modelización de distintos polinomios (Rogosa, 1995; Wainer, 2004).

II.3. Procedimientos de comparación de puntuaciones

Las grandes evaluaciones educativas basadas en los modelos de valor añadido, se caracterízan por el uso de diseños longitudinales que permiten medir el cambio que se produce en el constructo a estudiar a lo largo del tiempo. Para poder comparar las puntuaciones de las diferentes aplicaciones, es necesario un sistema métrico que permita la comparación de los resultados procedentes de un mismo instrumento o de instrumentos distintos.

En este tipo de modelos donde los instrumentos miden distintos contenidos, las escalas verticales son la opción más recomendada (Goldschmidt, Choi & Martinez, 2004; Singer & Willett, 2003). Sin embargo, acorde a las características de las pruebas, existen otras alternativas como la equiparación o la predicción, que también permiten la comparación de las puntuaciones en los diseños longitudinales.

El interés por comparar las puntuaciones, tiene su origen en Estados Unidos en la década de los 80, donde se realizaba un uso masivo de los test con fines de acceso a las universidades, al servicio militar, promociones, etc. Algo que obligó a trabajar con formas distintas de un mismo test, y, por tanto, la necesidad de comparar las puntuaciones de manera fiable para no cometer injusticias que repercutieran en los individuos (Muñiz, 2003).

El interés surgido de esta necesidad dio lugar a una continua revisión de métodos que permitieran la comparación de resultados y a la explicación de los mismos por parte de los constructores de test, junto con los nuevos modelos emergentes de la psicometría (Teoría de la Respuestas al Ítem, TRI) que permitieron la superación de muchas de las limitaciones de la Teoría Clásica de los Test (TCT). Estas razones desencadenaron el interés por la comparación de las puntuaciones dentro del campo de la investigación (Muñiz, 2003).

Autores como Dorans, Pommerich y Holland (2007) consideran que dicho auge de los procesos de comparación, también se deben al desarrollo de las evaluaciones basadas en la rendición de cuentas (accountability) que pretendían desarrollar sistemas de comparación rigurosos, con la finalidad de dotar a los resultados obtenidos en dichas evaluaciones de una mayor claridad y austeridad.

Así, poco a poco el estudio de los distintos procedimientos que permitían la comparación de las puntuaciones se fue convirtiendo en un tema de interés y de gran impacto, debido a la gran importancia que suponía y supone en la calidad de las evaluaciones educativas. En este sentido, los procedimientos comparativos son esenciales en cualquier proceso de evaluación que produce nuevas ediciones a lo largo del tiempo (Dorans, Moses & Eignor 2011).

II.3.1. Predicción, equiparación y escalamiento

Tal y como se ha señalado en el inicio de este apartado II.3, los procesos de comparación de puntuaciones dependen de las características de las pruebas que se pretenden comparar. Holland y Dorans, (2006) y Holland (2007) han descrito los mismos en profundidad mediante el establecimiento de una taxonomía basándose en la validez, la precisión y la reversibilidad con que se modifican las puntuaciones. Así, se pueden distinguir fundamentalmente tres procedimientos: predicción (*prediction*), equiparación (*equating*) y escalamiento (*scaling*). En este trabajo se prestará una especial atención en el escalamiento vertical, ya que es el procedimiento más utilizado, para comparar los resultados de los modelos de valor añadido a lo largo del tiempo.

A) Predicción

Según Holland, Dorans y Petersen (2007) la predicción es el procedimiento más antiguo utilizado para la comparación de las puntuaciones. De acuerdo con Dorans (2000) de los tres tipos de procedimientos mencionados anteriormente es el menos restrictivo y exigente.

A través del mismo, lo que se pretende básicamente es predecir los resultados esperados de una evaluación, a partir de otra información que puede proceder de variables socio-demográficas que guardan relación con el constructo a medir, o bien, que provienen de otras evaluaciones (Holland & Dorans, 2006). Un ejemplo de este procedimiento, seria intentar predecir los resultados de la prueba de acceso a la universidad, a partir de los resultados obtenidos en bachillerato.

Mediante la predicción se pretende pronosticar una puntuación a partir de otra puntuación o puntuaciones. En este sentido, a nivel metodológico hay que destacar la asimetría existente entre la variable predicha y las predictoras, en definitiva, la predicción tiene una única dirección (Holland & Dorans, 2006). Esto se debe a la utilización de la regresión lineal, donde la recta de regresión para predecir Y, en función de X, no es la inversa de la recta para predecir X, en función de Y (Galton, 1988).

Para predecir las puntuaciones de una prueba a partir de otra se utiliza una población que contesta a ambas pruebas. Una vez que están contestadas las dos pruebas, se trata de predecir la distribución condicional de la prueba B, para cada puntuación en la prueba A, mediante métodos de regresión lineal (Holland, 2007). Esto supone que lo que se predice son distribuciones de puntuaciones, no valores individuales.

También es posible predecir la puntuación verdadera, por ejemplo, Holland y Hoskens (2003) analizan la puntuación verdadera de un a prueba, a partir de la puntuación observada de la misma. Del mismo modo es posible la predicción a partir de la puntuación observada de otra prueba, que no había sido elaborada con la finalidad de ser escalada con la primera, de esta manera se habla de la predicción directa de la puntuación verdadera y predicción indirecta de la puntuación verdadera (Holland & Hoskens, 2003)

La predicción ha sido uno de los métodos más primitivos para la comparación de puntuaciones, pero no es un método óptimo para realizar comparaciones, tal y como evidencian muchos autores (Angoff, 1971; Holland & Dorans, 2006; Linn, 1993; Mislevy, 1992).

B) Equiparación.

La equiparación (*equating*, en inglés) hace referencia a las técnicas utilizadas para desarrollar instrumentos de evaluación, cuyos resultados se puedan intercambiar. La equiparación es el único proceso que permite que las pruebas sean estrictamente comparables e intercambiables (Kolen & Brennan, 2004), ya que el objetivo de este proceso es conseguir la equivalencia entre los tests a equiparar. De esta manera, las puntuaciones de un instrumento vienen expresadas en los términos del otro, independientemente de la forma a la cual el sujeto haya dado respuesta, asegurando una medida válida de la habilidad subyacente del individuo evaluado (Von Davier, 2011).

Dichas formas deben construirse intentando mantener tanto el contenido de las mismas como las propiedades psicométricas (Kolen & Brennan, 2014). Aunque, a pesar de dichos esfuerzos, generalmente suele existir diferencias en la dificultad, fiabilidad o contenido. La equiparación ajusta las diferencias en dificultad, pero no en contenido (Kolen & Brennan, 2014) para ello es necesario el uso de otros procedimientos como el escalamiento o *scaling,* que permite comparar los resultados de los estudiantes en distintos niveles educativos.

Según Kolen y Brennan (2004) la equiparación es el procedimiento estadístico que se usa para ajustar las puntuaciones en distintas formas de un test de modo que sean intercambiables. Ajusta diferencias en dificultad, entre formas que se han construido para ser similares en contenido.

En definitiva, el objetivo fundamental de la equiparación es poner en una escala común los resultados de uno o más instrumentos, que evalúan el mismo constructo, que tienen la misma fiabilidad, y que cuentan con especificaciones similares, de tal manera que dichas pruebas sean intercambiables.

Algunas situaciones en las que es posible el uso de la equiparación son, por ejemplo:

1) Contextos en los que los resultados de las pruebas tienen impacto para las puntuaciones de los evaluados (por ejemplo, pruebas de selección universitaria, evaluación académica, certificación laboral o de acreditación de competencias). Se busca minimizar los efectos de la copia entre los evaluados y que las puntuaciones en las distintas formas sean comparables.
2) Contextos en los que es necesario cambiar/actualizar las pruebas (por cambios curriculares, seguridad de las pruebas, publicación de ítems, etc.) y se busca que los resultados sean comparables de un año al siguiente (por ejemplo, para evaluar cuánto ha mejorado el rendimiento académico en una comunidad autónoma).
3) Estudios con diseño pre-test y post-test, en los que se mide el estado inicial de una variable antes de una intervención (escolar, clínica, etc.) y luego se mide el estado final de la misma variable, post intervención.
4) Creación o actualización de bancos de Ítems de un Test Adaptativo Informatizado (TAI). Por ejemplo, en la creación de un banco de 200 ítems se requiere distribuir los ítems por cuadernillos, puesto que no es factible que cada evaluado responda a todos los ítems.

La equiparación a nivel metodológico se considera un método robusto, pero al mismo tiempo requiere de muchos requisitos estadísticos y supuestos:

a) Simetría: la función utilizada para transformar las puntuaciones de la Forma X de un test, en una puntuación de la Forma Y, debe ser la inversa de la utilizada para transformar una puntuación de la Forma Y, en una puntuación de la Forma X.
b) Tests con las mismas especificaciones: las formas X e Y deben haber sido construidas con el mismo contenido (áreas a medir, tipo de ítems, número de ítems por área de contenido), y especificaciones estadísticas (media y desviación típica del test, distribución de los índices de dificultad y discriminación de los ítems).
c) Equidad: las medias de las distribuciones condicionales de las puntuaciones equiparadas en cada forma, respecto a las puntuaciones verdaderas, deben ser idénticas.
d) Invarianza poblacional: las transformaciones realizadas como consecuencia

de la equiparación serán las mismas independientemente de la población.

Dentro de la equiparación es posible distinguir entre equiparación vertical y equiparación horizontal. La equiparación horizontal se lleva acabo cuando las pruebas miden el mismo constructo y cuentan con el mismo nivel de dificultad. Cuando la dificultad es distinta, se utiliza el procedimiento de la equiparación vertical (Skaggs & Lissitz, 1982)

El ajuste de las diferencias en dificultad que existen en las distintas pruebas, que han sido diseñadas para tener el mismo contenido y dificultad, es denominado estrictamente equiparación por Kolen y Brennan (2004), ya que es el único procedimiento que permite el establecimiento de puntuaciones que son realmente intercambiables.

Según Skaggs y Lissitz (1982) en la equiparación horizontal se cuenta con instrumentos interconectados que miden un mismo constructo, con las mismas características psicométricas en una misma población. Mientras que, en la equiparación vertical los instrumentos se diseñan para evaluar el mismo constructo, pero con diferentes niveles de dificultad, siendo aquí el propósito la equiparación de pruebas que miden el mismo rasgo, pero dentro de un amplio rango de dificultad. Estos mismos autores consideran que la equiparación vertical es más complicada que la horizontal, ya que mediante dicho procedimiento se pretende establecer puntuaciones que aunan la misma dimensión en diferentes grados, cuando las pruebas se diseñan con diferente dificultad, puesto que la intención es medir grupos con diferentes niveles de habilidad.

A veces estas formas de equiparación pueden aparecer simultáneamente con otros procedimientos de comparación. Por ejemplo, se utiliza la equiparación horizontal cuando la prueba se realiza en cada uno de los cursos a evaluar, utilizando más de una misma forma para el mismo curso académico, pero si además se pretenden comparar las puntuaciones en distintos grados (diferencias de contenido) es necesario procedimientos de escalamiento vertical.

Esta mezcla de procedimientos con el fin de conseguir la comparabilidad de las puntuaciones, a veces hace que sea difuso distinguir entre equiparación y otros términos afines como puede ser el escalamiento. Pero actualmente se ha reservado el término equiparación de manera exclusiva para los casos en los que las

puntuaciones sean intercambiables, en pruebas que cuenten con las mismas especificaciones para medir el mismo constructo y donde se cumpla los principios de simetría e invarianza (Éxposito, 2015)

C) Escalamiento

Los métodos actuales de escalamiento denominados *scale aligning,* pretenden proporcionar puntuaciones que sean comparables mediante la transformación de los resultados de pruebas que son diferentes, a una misma escala (Gempp, 2010).

La consecución de puntuaciones que sean comparables, se realiza de manera indirecta conectando los dos test de forma externa mediante un tercer test, o un test de anclaje que puede haber sido creado con otros fines ajenos al escalamiento de las pruebas (Holland, 2007).

Existen diversos tipos y subtipos de escalamiento. Holland y Dorans (2006) y Holland (2007) se basan en el propósito, diseño y características técnicas de las pruebas para clasificarlos. De esta manera distinguen entre escalamiento de baterías (*battery scaling*) propuesto por Kolen (2004), escalamiento de anclaje (*anchor scaling*) que incluye el escalamiento a una población hipotética (*scaling on a hipothetical population*) y el escalamiento vertical (*vertical scaling*) propuesto por Kolen y Brennan (2004), calibración (*calibration*) y concordancia (*concordance*) propuesto por Dorans (2004). A continuación, se presentan las características más importantes de cada uno de ellos.

1. Escalamiento de baterías.

Dicho procedimiento permite transformar las escalas de cada test con el fin de que las distribuciones sean comparables. Este procedimiento denominado *test battery scaling* por Kolen (2004), se utiliza en pruebas que miden diferentes constructos pero que se aplican a una misma población. Se caracteriza porque permite el establecimiento de escalas comparables, pero no de puntuaciones equivalentes.

La recogida de datos en dicho enfoque se puede realizar de tal manera que una misma muestra de examinados realicen las dos formas completas, o bien, a partir de muestras equivalentes de sujetos pertenecientes a una población común

utilizando uno de los test realizados. En cualquier caso, los test son respondido por grupos que son equivalentes a la población de referencia (Holland, 2007).

2. Escalamiento de anclaje

Este procedimiento denominado "Escalamiento mediante medidas de anclaje" (*anchor measure*), permite comparar los resultados de pruebas que miden constructos diferentes en poblaciones distintas. Para ello, se utiliza una tercera prueba que se aplica a las dos poblaciones, y que correlaciona con los constructos evaluados (Holland, 2007). Dicha prueba sirve de medida de anclaje, de tal manera, que permite situar las puntuaciones en una escala común.

3. Escalamiento a una población hipotética

Dentro del escalamiento de anclaje, una de las aproximaciones estadísticas que supone un uso diferenciado de las medidas de anclaje, es el escalamiento a una población hipotética (*Scaling on a Hipothetical Population).* En este procedimiento se proyecta las distribuciones que corresponderían a cada prueba, a partir de los resultados obtenidos en la medida de anclaje, por ejemplo, suponga que Y denota una prueba a escala y A es la otra medida. La fecha en que los exámenes Y y A son realizados, se usan para estimar la distribución condicional de Y dada A en la población de examinados, pudiendo ser quienes realizan el test Y, distintos de los que realizan la prueba A (ver Holland & Dorans, 2006).

4. Calibración

La calibración se aplica a distintas medidas de un mismo constructo. Las medidas difieren en fiabilidad (forma larga y corta de un test), por ejemplo, cuando el test corto es una versión de screening del test más largo.

Aunque a veces dicho término de calibración no es claro y se mezcla con el de escalamiento vertical, por ejemplo, Angoff, (1971) considera que la calibración es un método apropiado cuando se pretende comparar los resultados de dos pruebas que se han definido para medir el mismo constructo, pero en niveles educativos distintos, o que tienen diferente fiabilidad. En este caso, el término adecuado es calibración en vez de equiparación vertical, ya que este último procedimiento no se podría llevar acabo, porque no se cumplen algunos de los criterios necesarios para el mismo.

Otros autores hablan de calibración, aludiendo al proceso mediante el cual se obtienen los parámetros de los ítems analizados utilizando procedimientos de TRI (Lord, 1980).

Kolen y Brennan (2014) utilizan el término calibración en el mismo sentido que Dorans y Holland (2006) y Holland (2007), ya que consideran que es un procedimiento en el cual los instrumentos miden el mismo constructo, con niveles de dificultad similares, pero existen diferencias en la habilidad o incluso en la longitud de las pruebas.

Dentro de la TRI se pueden distinguir tres procedimientos de calibración, con el objetivo de situar las estimaciones de diferentes tests en una escala común. Dichos procesos de calibración se realizan de forma similar, tanto en la equiparación horizontal como, en el escalamiento vertical.

Calibración por separado (CS): mediante este procedimiento los parámetros de los ítems en las diferentes pruebas, se estiman por separado realizando posteriormente una transformación lineal de los mismos. Dicha transformación puede realizarse de diferente manera, mediante métodos basados en los momentos como son el método de media/media (Loyd & Hoover, 1980) y media/sigma (Marco, 1977) o a través de métodos basados en las curvas características del ítem de Haebara (1980) y Stocking y Lord (1983).

En el caso del método media/media las medias de a y b tienen que ser iguales después de la equiparación:

Ec.II.1

$$A = \frac{\mu(a_X)}{\mu(a_Y)}$$

$$B = \mu(b_Y) - A\mu(b_X)$$

Los μ se refieren a los parámetros de discriminación y dificultad de a y b en ambas escalas (pruebas X e Y).

En el método de media/sigma las medias y varianzas de b tienen que ser iguales después de la equiparación.

Ec.II.2

$$A = \frac{\sigma(b_Y)}{\sigma(b_X)}$$

$$B = \mu(b_Y) - A\mu(b_X)$$

Los σ se refieren a las desviaciones típicas de los parámetros de dificultad de los ítems, en las pruebas X e Y, y μ son las medias de esos mismos parámetros.

Dichos métodos basados en los momentos, no consideran todos los parámetros conjuntamente. Aun así, de ambos métodos Ogasawara (2002) considera que el método media/media es más estable que el media/sigma.

Los métodos basados en la curva característica, determinan los parámetros de la transformación que hacen lo más parecidas posibles las Curvas Características de los Ítems (CCIs) de anclaje entre ambos tests. Para ello, se basan en una relación lineal centrada en la igualdad, que se encuentra vinculada a la probabilidad de responder correctamente a un ítem de acuerdo a sus parámetros.

El método de Stocking-Lord, se basa en el cálculo de las diferencias cuadráticas entre las curvas características de un test, para un valor concreto del rasgo θ_i que es el sumatorio de las curvas características de los ítems comunes entre formas.

Ec.II.3

$$SLdif(\theta_i) = \left[\sum_{j=1}^{V} P_{ij}\left(\theta_{Yi}\middle|\hat{a}_{Yj}, \hat{b}_{Yj}, \hat{c}_{Yj}\right) - \sum_{j=1}^{V} P_{ij}\left(A\theta_{Xi} + B; \middle|\frac{\hat{a}_{Xj}}{A}, A\hat{b}_{Xj} + B, \hat{c}_{Xj}\right)\right]$$

Cuando ya se han calculado las diferencias, se acumulan mediante la utilización de todas las puntuaciones de los individuos. Los parámetros A y B se obtienen cuando el proceso iterativo encuentra los valores que minimizan el siguiente criterio (Navarro, 2013)

Ec.II.4

$$SLCrit = \sum_{i} SLdif(\theta_i)$$

En el caso del método Haerbara, se basa en la utilización de la suma de las diferencias cuadradas entre las curvas características de cada reactivo común para individuos con un determinado nivel de rasgo (θ_i) (Navarro, 2013), la suma de esas diferencias es:

Ec.II.5

$$\mathrm{Hdif}(\theta_i) = \sum_{j=1}^{V} \left[P_{ij}\left(\theta_{Yi} \middle| \hat{a}_{Yj}, \hat{b}_{Yj}, \hat{c}_{Yj}\right) - P_{ij}\left(A\theta_{Xi} + B; \left| \frac{\hat{a}_{Xj}}{A}, A\hat{b}_{Xj} + B, \hat{c}_{Xj}\right)\right]^2$$

Aquí, las diferencias se suman con los resultados de todos los casos, y la obtención de las constantes A y B son similares al método anterior (Navarro, 2013).

En general se ha encontrado que los métodos basados en la curva característica producen resultados más estables que los basados en los momentos (Ogasawara, 2002)

Calibración Conjunta (CC): dicha calibración se caracteriza porque los parámetros de los ítems de las distintas formas del test, se realizan de manera simultánea mediante un modelo multigrupo. Para ello, las respuestas de los individuos en cada aplicación se tratan como un grupo distinto y aquellos ítems que no han sido contestados por todos los individuos se consideran como valores perdidos.

Este procedimiento permite establecer una escala común a la hora de estimar los parámetros de los ítems, en grupos de sujetos que responden a diferentes test, es decir, las respuestas de distintos grupos, que contestan a diferentes formas del test, pero que cuentan con ítems comunes, pueden calibrarse a la vez como si se tratara de una misma prueba que ha sido respondida por todos los sujetos.

Calibración Fija (CF): en este caso los parámetros de los ítems que son comunes se mantienen invariantes entre las aplicaciones. En primer lugar, se estiman los parámetros de los reactivos que van a ser la base de la escala de manera separada. Para fijar esa escala en la prueba siguiente, se usan los valores estimados para los ítems comunes (ítems de anclaje). Los parámetros de dichos ítems permanecen fijos mientras que los específicos de la nueva forma se calibran.

De los diferentes procedimientos de calibración, Kolen y Brennan (2004) señalan que la calibración conjunta es quizás la más adecuada para el escalamiento vertical, ya que al realizar una única estimación utilizan toda la información disponible de los parámetros de los ítems, mientras que en la calibración por separado es necesario además del uso de varias transformaciones para situarlo en una escala común, el uso de varias estimaciones. Sin embargo, a nivel práctico la calibración por separado puede ser más idónea, ya que permite la comparación de las estimaciones de los parámetros aplicación tras aplicación, y, por lo tanto, identificar ítems que se comportan de manera diferente entre dichas aplicaciones. Asimismo, la calibración conjunta puede dar lugar a problemas de convergencia, ya que se tiene que estimar un gran número de ítems en una sola vez, incluyendo los reactivos perdidos por el diseño.

Jungnam (2007) compara los tres procedimientos de calibración (CS, CC y CF) utilizando datos educativos basados en la resolución de problemas, y concluye que no existen diferencias de medias en los tests aplicados en cursos consecutivos al comparar los tres métodos de calibración. Pero es verdad que en los grados inferiores el método de calibración conjunta produce diferencias de medias menores, que el método de calibración por separado y calibración fija, entre los cuales las diferencias de medias son iguales. Respecto a la comparación de las desviaciones típicas producidas por los diferentes métodos son casi inexistentes, de manera similar ocurre con el tamaño del efecto, y las distancias horizontales de los diferentes tipos de calibración

En definitiva, la utilización de un tipo u otro procedimiento de calibración, dependen de la decisión que tome el investigador de acuerdo a sus necesidades.

5. **Concordancia:** este procedimiento es utilizado cuando se miden constructos similares, que se evalúan con pruebas que cuentan con la misma fiabilidad y dificultad en una misma población, pero han sido construido de acuerdo a diferentes especificaciones técnicas (Holland, 2007).

Dicho procedimiento tiene como resultado el establecimiento de una tabla de concordancia que establece la correspondencia entre ambas pruebas. Así, dos individuos que han realizado dos pruebas pueden comparar sus resultados. En el caso de que solo hayan respondido a una prueba, es posible predecir cuál hubiera

sido su resultado en la otra prueba.

II.3.2. Escalamiento vertical

Según Hollan y Doran (2006) y Holland (2007) este procedimiento es el más idóneo cuando se miden constructos similares que contienen distinta dificultad, por ejemplo, estudiantes de distintos grados.

Él término escalamiento vertical (*vertical scaling*) (Kolen & Brennan, 2004), también es denominado como "calibración de test en distintos niveles de habilidad" (Angoff, 1971), "escalamiento para la comparativa" (AERA, APA, NCME, 1985), "calibración" (Linn, 1993; Misslevy, 1992), "equiparación vertical" (Crocker & Algina, 1986; Skaggs & Lissitz, 1982), "enlace vertical" o "escalamiento transversal" (Carlson, 2011).

La aparición del escalamiento vertical se encuentra íntimamente relacionado con la aparición de las grandes evaluaciones educativas y con la rendición de cuentas (Dorans, et al., 2007). Es, sobre todo, a partir de la mencionada ley *No Child Left Behind* (2001) cuando se plantea la necesidad de comparar los resultados de los estudiantes a lo largo del tiempo de manera fiable, ya que en dichas evaluaciones se estudiaban las puntuaciones de los sujetos desde el tercer al octavo grado. De tal manera que las puntuaciones entre los distintos cursos no eran directamente comparables por el simple hecho de estar ancladas en un año específico.

En este contexto aparece el escalamiento vertical con el objetivo de situar en una escala común los resultados de los cursos evaluados, de manera que fuera posible comparar las puntuaciones de los estudiantes a lo largo del tiempo.

Las evaluaciones deben estar escaladas verticalmente y a su vez pueden estar equiparadas horizontalmente, si las puntuaciones de escala se basan en diferentes formas de una evaluación (diseñadas para un grado dado o nivel de logro) que se califican en la misma escala de desarrollo. Todos los exámenes estatales deben ser equiparados horizontalmente, para que sea posible determinar si las tasas de competencia escolar y estatal han cambiado con el tiempo.

La elaboración de dichas escalas verticales es un procedimiento fundamental dentro de las evaluaciones de valor añadido, basadas en modelos de crecimiento y ganancia, porque los efectos de las escuelas dependen en gran medida de cómo se

ha construido la escala (Briggs & Betebenner, 2009; Briggs & Weeks, 2009; Briggs, Weeks & Wiley, 2008; Jungnam, 2007). Aun así, no todos los MVA utilizan dicho procedimiento debido a que o bien tan solo se basan en dos puntuaciones para el desarrollo de un modelo de regresión multinivel, donde el rendimiento previo se utiliza como covariable (Demie, 2003; Ray et al., 2009) o bien, porque no son un requisito necesario, tal y como por ejemplo ocurre en el modelo de Tennesse (Sanders & Horn, 1994; Sanders et al., 1997)

Cuando no se utiliza una escala vertical para la comparación de puntuaciones, los expertos de pruebas utilizan técnicas estadísticas para aproximar el cambio en el crecimiento de año en año (Gong, Perie & Dunn 2006). A falta de una escala vertical, los datos se suelen convertir en una escala normalizada (McCall et al., 2004) para comparar el rendimiento de los alumnos entre sí. La conversión a una escala normal supone que alguien siempre estará por encima o por debajo del promedio, con independencia de que los estudiantes estén cumpliendo con las normas o no. De esta manera, los políticos y los educadores deben plantearse si un enfoque de escala normalizada es lo más idóneo en un sistema basado en estándares educativos.

La razón principal para desarrollar una escala vertical en los MVA, es la necesidad de comparar en las evaluaciones educativas los niveles de logro a lo largo del tiempo (Harris, 2007) de manera segura y fiable, ya que en las mismas se encuentran implicados una gran variedad de agentes, como son las familias, estudiantes, docentes, instituciones, administraciones, etc.

Si no se garantiza la comparación de las puntuaciones en el estudio del cambio o del crecimiento, en los procesos de aprendizaje a lo largo del tiempo, las conclusiones extraídas a partir de los resultados obtenidos pueden conllevar la toma de decisiones erróneas.

II.3.2.1. Característica y problemas asociados al escalamiento vertical

En teoría, la métrica óptima para examinar el cambio producido en las puntuaciones d los estudiantes, es una escala equiparada verticalmente basada en la TRI, porque está en una escala de intervalo y es comparable entre los distintos grados (Hambleton & Swaminathan, 1987).

Las escalas verticales se elaboran a partir de distintas pruebas, adecuadas a los estudiantes de un determinado curso. Dicho procedimiento pone en la misma

escala los resultados de varios instrumentos que miden un constructo similar, pero en niveles educativos diferentes, por lo que cambian en dificultad y en contenido, aunque generalmente el contenido suele ser afín.

Mediante dichas escalas es posible conocer si un estudiante aumenta su conocimiento en el constructo evaluado, ya que si su puntuación en la escala se ve incrementada significa que su nivel de habilidad ha crecido (Lissitz, Doran, Schafer & Willhoft, 2006).

Dichas escalas cuentan con una serie de propiedades que se detallan a continuación (Navarro, 2013):

A) Característica de intervalo

Esta propiedad supone que la distancia entre distintos puntos es la misma, es decir, entre la puntuación 3 y 6 hay la misma distancia que entre 6 y 9. Esto a nivel práctico supone que los diferentes puntos de la escala significan lo mismo, un tema un tanto controvertido (Ballou, 2009), ya que dicha asumción solo puede ser cierta si el modelo psicométrico utilizado para la construcción de la escala, supone una representación matemática verdadera entre el nivel de rendimiento del estudiante y sus respuestas a los ítems (Martineau, 2009).

Así la problemática reside en que, aunque el uso de escalas verticalmente equiparadas permite comparar de manera cuantitativa el cambio que se produce en los aprendizajes en los diferentes grados evaluados, empleando la TRI, el uso de dicha metodología supone que la escala cuenta con la propiedad de intervalo; cuando existe cierto debate por parte de los expertos en psicometría sobre si realmente dichas escalas cuentan con esta propiedad (Ballou, 2009; Briggs & Betebenner, 2009).

Mediante los modelos psicométricos, es posible representar la probabilidad que un estudiante tiene de responder correctamente a un ítem, acorde a su nivel de logro. De esta manera, los ítems con un buen funcionamiento deberían tener una alta probabilidad de ser respondidos correctamente, por alumnos que cuentan con altos niveles de rendimiento. Sin embargo, la relación entre la probabilidad de responder correctamente a un ítem y el nivel de logro no es clara.

Algunos autores consideran que no es posible validar las condiciones necesarias para que estas escalas cuenten con la propiedad de intervalo (Briggs et al., 2008), y, por lo tanto, solo se pueden considerar como escalas ordinales, lo que supone el uso de una información que es menos precisa, y no es una solución para las medidas de resultados del VA (Thum, 2006).

Una opinión contraria, es la expuesta por Reckase (2008) quien considera que las escalas verticales construidas bajo los supuestos de la TRI, si cuentan con la propiedad de intervalo por dos motivos. El primero de ellos, es que dicho autor considera que si el modelo de TRI utilizado ajusta con los datos si se puede hablar de escala de intervalo, ya que la forma de la función dentro de la TRI no está definida a menos que la escala de rendimiento tenga la propiedad de intervalo. El segundo motivo es que si se considera que la distribución de las puntuaciones verdaderas y las observadas en el test es coincidente con la distribución asumida, se puede decir que los resultados cuentan con la propiedad de intervalo. Además, normalmente para referirse al rendimiento de los estudiantes se asume una distribución normal, así, si el número de respuestas correctas sigue dicha distribución puede concluirse que la escala cuenta con la característica de intervalo.

B) Dimensionalidad del constructo evaluado

Uno de los objetivos de las escalas verticales es medir un nivel de rasgo a lo largo del tiempo, esto puede ser un problema, ya que, si el rasgo medido cambia de manera significativa a lo largo de los cursos académicos, puede suponer la violación del supuesto de unidimensionalidad del constructo (supuesto fundamental de la TRI), y, por lo tanto, afectar a la escala y a las estimaciones finales del VA.

En un estudio de Lockwood et al. (2007) en el que se analiza la sensibilidad de los MVA ante la utilización de dos subescalas de rendimiento, muestran que las estimaciones de los MVA se ven más afectadas por la variación de la escala, que por los cambios que se producen en los modelos estadísticos, debido a la introducción de predictores contextuales, la utilización de análisis de ganancia, ajuste de covariables, etc.

Por lo tanto, los diseños de VA añadido pueden verse afectados por los cambios de etapa o ciclo educativo, ya que, los mismos pueden influir en la

estructura de la escala de las puntuaciones de logro y, en definitiva, al diseño de los MVA (Gaviria, Biencinto & Navarro, 2009).

En este sentido quizás es mejor centrarse en una etapa o ciclo educativo concreto que en un gran número de cursos, ya que, las diferencias en los contenidos curriculares pueden afectar a las escalas verticales, debido a que la dimensionalidad del constructo va siendo cada vez menos uniforme, a medida que los cursos evaluados son superiores.

Martineau (2009) considera que el cambio que se produce en el constructo evaluado a lo largo del tiempo, puede afectar a las estimaciones cuando se utilizan las escalas verticales en el análisis del VA. De tal manera que, una escala vertical no puede ser utilizada de manera válida en las evaluaciones de alto impacto para estimar el valor añadido. Este mismo autor añade que el uso de puntuaciones de escala multidimensionales, pueden no tener utilidad de carácter práctico en la medida del VA de las escuelas o docentes. Por ello recomienda para estimar el VA, medir un único constructo teniendo en cuenta el curso actual y el previo. Denominando esta aproximación como análisis de Valor Añadido en una pareja de cursos unidimensionales. También, recomienda como otra posible opción para solventar el problema de la dimensionalidad, la utilización de métodos de escalamiento vertical multidimensionales basados en la TRI.

En definitiva, la falta de unidimensionalidad hará que la validez de las inferencias basadas en una escala vertical, se vea disminuidas según avanza la distancia entre los cursos evaluados.

C) Varianza del crecimiento a lo largo del tiempo

Ballou (2009) considera que las medias de las ganancias de los cursos superiores y la varianza del rendimiento en una escala vertical, disminuyen al utilizarse la TRI. De esta manera se plantea si los estudiantes que están en los cursos iniciales, aprenden más que los que se encuentran en los cursos superiores.

Es evidente que este tipo de escalas suponen que la ganancia varía de una parte a otra de la escala, por ejemplo, los estudiantes que consiguen cinco puntos en la parte alta de la escala, no consiguen la misma ganancia que los estudiantes que

consiguen esa misma cantidad de puntos, pero en la parte baja de la escala (Patz, 2007).

En este sentido, es menos problemático comparar las estimaciones de dos cursos consecutivos que pertenecen a una misma cohorte de estudiantes, que comparar los datos de estudiantes que pertenecen a cursos muy distantes, por ejemplo, puntuaciones de alumnos de 2º de educación primaria con puntuaciones de estudiantes de 3º de educación secundaria, ya que, a nivel de escala, los alumnos que se encuentran en la parte alta de la misma se pueden ver afectados por el conocido *efecto techo* donde el nivel de logro de estos estudiantes se ve disminuido, debido a que se aproximan a las puntuaciones máximas de la escala. Mientras que los estudiantes que se encuentran en la parte baja, se pueden ver afectados por el *efecto suelo* caracterizado porque los estudiantes que se sitúan al comienzo de la escala experimentan un mayor crecimiento (Betebenner & Linn, 2010). Ambos efectos, techo y suelo, pueden ser problemático para analizar el crecimiento que se produce entre cursos, pero no dentro de un mismo curso.

Estos fenómenos se encuentran íntimamente relacionado con la característica de intervalo con la que se quiere dotar a las escalas verticales, ya que, si dichos efectos aparecen, los intervalos entre las puntuaciones de escala no son iguales, y, por lo tanto, no es lo mismo crecer cinco puntos al principio de la escala que al final. En este sentido Reckase (2010) considera que el uso de escalas elaboradas con TRI, como no tienen límites reducen estos efectos. Por otro lado, Braun et al. (2010) creen que la escala de intervalo es una forma de medir que puede o no, coincidir con el valor que de manera práctica se le da a las diferencias entre intervalos.

Todas estas propiedades de las escalas verticales pueden ser un punto crítico en la estimación de los MVA, que utilizan las mismas como variable de resultados, ya que las estimaciones del VA depende de cómo se haya construido la escala (Briggs & Betebenner, 2009; Lockwood et al., 2007; Ray, 2006).

Por lo tanto, si se utiliza este tipo de escala hay que ser consciente de cómo la misma puede afectar a los MVA. Generalmente, el uso de la TRI es la opción más común (debido a su fiabilidad y validez) para el procedimiento de escalamiento vertical con el fin de conseguir la comparabilidad de las puntuaciones. El uso de la

TRI conlleva una serie de problemas estudiados anteriormente que no deben pasar desapercibidos, especialmente cuando las escalas verticales se usan con la finalidad de tomar decisiones sobre los resultados de los sistemas educativos basados en la rendición de cuentas.

Como opinión personal, el escalamiento vertical es la opción más idónea cuando el número de cursos que se quieren anclar no es muy elevado, siendo especialmente exitoso cuando se comparan cursos adyacentes, ya que, la escala empieza a producir problemas al aumentar las distancias entre los grados a estudiar.

II.3.2.2. Elaboración de una escala vertical

En el diseño de evaluaciones basadas en el Valor Añadido, la elaboración de una escala vertical es imprescindible cuando se utilizan modelos longitudinales con pendiente de crecimiento o bien modelos que utilizan una puntuación de ganancia como variable de resultados. En estos casos con el fin de estudiar el cambio que se produce a lo largo del tiempo, es necesario contar con puntuaciones que se encuentren en una escala común para que se puedan comparar a lo largo del tiempo.

Por lo tanto, la elaboración adecuada de una escala vertical es fundamental, ya que, de ella depende la obtención de estimaciones fiables y válidas. Además, a partir de las puntuaciones de la escala vertical, se toman una serie de decisiones que pueden a afectar a todos los agentes vinculados con la rendición de cuentas.

En este proceso de elaboración de una escala, es necesario tener en cuenta una serie de aspectos claves (Navarro, 2013), que se explican a continuación:

A) Tipo de diseño para la recogida de datos

El primer aspecto a tener en cuenta, es el diseño que se considera más adecuado para recoger información como parte del proceso de equiparación.

Kolen y Brenan (2004) distinguen entre tres tipos de diseño que son los más utilizados en el proceso de anclaje vertical, estos son, el diseño de ítems comunes, diseño de grupos equivalentes y diseño de test de anclaje. De estos tres, la inclusión de ítems de anclaje entre los tests de los diferentes grados estudiados, es el diseño más utilizado (Briggs & Weeks, 2009; Chin, Kim & Nering, 2006; Jungnam, 2007; Tong & Kolen, 2007).

Además de los ítems de anclaje, también es posible utilizar un test de anclaje como en el estudio de Lee y Ban (2010) donde en la primera aplicación se administra una única forma (Forma A), que es la base de la escala y en la siguiente toma de datos se administra una nueva forma (Forma B) además de la forma A, ambas formas distribuidas en un diseño en espiral. En este proceso espiral, se alternan las diferentes formas de los tests en los paquetes que se mandan a las escuelas, de tal manera que en algunos casos se asigna la forma A a la primera aula del listado y la forma B a la siguiente, y así de manera consecutiva. En otras aulas, sin embargo, se comienza al revés. Mediante este reparto aleatorio de cuadernillos se consiguen grupos equivalentes, donde la forma en que se administra ambas aplicaciones actúa de elemento de anclaje.

Tong y Kolen (2007) comparan los resultados entre un diseño con ítems de anclaje con un diseño con test de anclaje y concluyen que el diseño con ítems de anclaje muestra un mayor crecimiento, pero ambos tienden a reducir ese crecimiento a medida que el curso evaluado es superior.

Según Kolen y Brennan (2004) cuanto mayor número de ítems de anclaje existan, menor será el error de equiparación. De esta manera sugieren que al menos el 20% de los ítems de un test sean comunes en un diseño de equiparación horizontal.

Aun así, no existe mucha investigación sobre el número de ítems comunes que debe haber en un proceso de equiparación vertical, pues, por ejemplo, Patz (2007) recomienda el uso de 15 ítems de anclaje, mientras que Chin et al., (2006) considera que el número de ítems de anclaje, debe ser aquel que cubra el continuo de la habilidad estimada y permita disminuir el sesgo y los errores. Dichos autores demuestran que independientemente del número de ítems de anclaje, si el rango de dificultad de estos reactivos, varía hasta dos desviaciones típicas en las puntuaciones de rasgo, es decir, cubren la mayor parte del continuo de las puntuaciones de rasgo, las estimaciones mejoran, ya que se disminuye el sesgo puesto que los niveles del error de estimación son muy bajos.

B) Modelos psicométricos

Otro aspecto importante es el modelo psicométrico que se va a utilizar, siendo los modelos logísticos de la TRI los más utilizados.

Los procesos de equiparación y de anclaje se encuentran íntimamente relacionados con el escalamiento, ya que permiten asociar números u otros aspectos con los niveles de logro de los estudiantes, de tal manera que se establezca una escala común de las puntuaciones (Kolen & Brennan, 2004). Generalmente, dicha escala se crea a partir de las respuestas que los sujetos dan a un solo test, pero si se emplean diversas pruebas es necesario una transformación de las puntuaciones para compararlas. Esta transformación se puede realizar de diferentes maneras, pero lo normal es que se basen en procedimientos de la TRI, pues es la metodología más empleada para la construcción de escalas verticales basadas en los modelos de valor añadido (Briggs et al., 2008).

En los modelos TRI las respuestas de los sujetos a los ítems, dependen de las características de los reactivos y del nivel de rasgo que tienen los sujetos. Así, se estima una función de probabilidad de respuesta a cada uno de los reactivos que forman parte de una prueba, de acuerdo a los parametros de cada ítem y la puntuación que obtiene el individuo (Navarro, 2013). La TRI se centra en las propiedades individuales de los ítems, y permite conocer la precisión del test para cada nivel de rasgo. Kolen y Brenan (2004) denominan calibración al proceso que se utiliza para calcular los parámetros de los ítems de un test .

En el caso de ítems dicotómicos, existen modelos logísticos de uno, dos y tres parámetros que estiman la probabilidad que el individuo tiene de responder correctamente a un ítem.

El modelo logístico de un parámetro (ML1P) asume que los ítems varían sólo en el parámetro de dificultad, al que se denomina parámetro b. A mayor valor de b, en este modelo, menor será la probabilidad de acierto. Dicho parámetro de dificultad, indica el nivel de rasgo en que la probabilidad de acertar el ítem es del 50%.

El modelo de dos parámetros (ML2P) se diferencia del anterior, en que en este modelo se añade un segundo parámetro que ya no es constante, la discriminación (parámetro a). Este modelo permite que el parámetro a, que indica la capacidad discriminativa del ítem, varié de ítem a ítem. La discriminación indica en mayor o menor medida la inclinación o la pendiente CCI y es proporcional al ángulo que forma la CCI en relación al eje de abscisas. El valor del parámetro de discriminación oscila entre 0,3 y 2,5, considerándose ítems discriminativos los que tienen valores mayores a 1.

Finalmente, el modelo logístico de tres parámetros (ML3P) añade a los parámetros a y b un tercero, c que representa la probabilidad de acertar el ítem por azar. Exactamente el parámetro c o pseudoazar, supone la probabilidad de acierto para personas con un nivel de rasgo extremadamente bajo. Si no hay omisiones, suele tomar un valor próximo al inverso del número de opciones de respuestas, si hay omisiones, cuanto mayor sea el número de personas que no responden al ítem, menor será el parámetro c. Al ser dicho parámetro una probabilidad, sus valores pueden oscilar entre 0 y 1, pero suele estar entre 0 y 0,5.

Hay algunos autores que reconocen también el modelo logístico de cuatro parámetros (ML4P), cuyo fin es explicar aquellos casos en que los sujetos que tienen un nivel de aptitud alto, no responden correctamente al ítem, para ello se añade un cuarto parámetro que es el parámetro de descuido del ítem (parámetro d).

Briggs et al. (2008) estudian el efecto que los modelos de la TRI de uno y dos parámetros tienen sobre la elaboración de una escala vertical. A las conclusiones que llegan, es que el ML1P produce puntuaciones medias y desviaciones típicas menores que ML3P. Respecto al crecimiento entre cursos, ambos modelos muestran un crecimiento mayor en grados inferiores y una mayor distancia entre grados superiores.

C) Proceso de escalamiento y calibración de los ítems.

Otro aspecto importante es el escalamiento y la calibración de los ítems, ya que al utilizar los modelos TRI para situar las distintas formas en una escala común mediante la equiparación horizontal o el anclaje vertical, es necesario la modificación de los parámetros de los ítems mediante la calibración. Los métodos

de calibración son los estudiados en el *apartado II.3.1:* calibración conjunta, calibración fija y/o la Calibración por Separado.

D) Calificación de los sujetos

Finalmente, tampoco se puede obviar la calificación de los sujetos (*scoring*), la estimación de los niveles de logro.

Generalmente se utilizan tres métodos de estimación, Máxima Verosimilitud (MV), Bayes o Esperada a Posteriori (EAP) y Bayes Modal o Máxima a Posteriori (MAP), de estos tres métodos el MV es el más utilizado. Este procedimiento utiliza la función de verosimilitud que tiene en cuenta el patrón de respuesta de los individuos. El mayor problema de la MV, es que no produce valores para aquellos sujetos que no tienen al menos una respuesta correcta o que han contestado bien a todos los ítems.

Respecto a los procedimientos de Esperada a Posteriori (EAP) y Máxima a Posteriori (MAP), a la hora de realizar las estimaciones se basan en la estadística bayesiana. Dichos procedimientos combinan información procedente de la función de verosimilitud, con supuestos acerca de la distribución de la habilidad en la población (Navarro, 2013). La distribución asumida es la previa, que al combinarse con la información de verosimilitud se construye una distribución ajustada denominada distribución a posteriori. La diferencia entre ambos enfoques (EAP y MAP) se encuentra en el parámetro que utilizan para la distribución a posteriori, así mientras que la EAP utiliza la media de la distribución, MAP emplea la moda.

El uso de estimadores bayesianos también son conocidos como BLUP, y al igual que ocurre con el EAP o el MAP, tienden a contraer la escala debido a que las estimaciones se encuentran contraídas en torno a la media de la población, tal y como se estudiará en el capítulo IV.

Respecto a que método es el más idóneo para la estimación de las puntuaciones de los sujetos, Jungnam (2007) considera que los métodos bayesianos introducen mayor sesgo en las estimaciones, sobre todo en los puntos extremos, mientras que el método de MV no produce sesgos cuando se trabaja con test largos, pero cuentan con el problema de que no calcula valores cuando los estudiantes tienen todas las respuestas correctas o bien no tienen ninguna.

Lee y Ban (2010) también estudian las diferencias entre MV y los métodos bayesianos en un estudio simulado. En dicho trabajo encuentran que los resultados de los métodos bayesianos (EAP y MAP) tienden aparecerse entre ellos, y que las varianzas de ambos, son inferiores a las estimadas con MV. Sin embargo, el sesgo es mayor cuando se utilizan estimadores bayesianos.

Dichos resultados se encuentran en sintonía con los resultados obtenidos por Tong y Kolen (2007), donde en un estudio realizado obtienen valores inaceptables con test cortos de 25 ítems, debido al efecto de encogimiento ("*shrunkage"*) de las puntuaciones. Por ello, dichos autores recomiendan el uso de MV, ya que producen menos error que los estimadores bayesianos, especialmente cuando las poblaciones a equiparar tienen una habilidad muy distinta, sobre todo en test cortos.

II.4. Variables contextuales

Otro aspecto importante que puede afectar a las estimaciones finales de los Modelos de Valor Añadido es la introducción de variables contextuales (Ballou et al., 2004; Choi et al., 2006; Ferrão, 2009; Lockwood et al., 2007; Tekwe et al., 2004). Esta cuestión se encuentra estrechamente relacionada con los modelos de crecimiento, ya que, algunos investigadores (Sanders & Horn, 1994; Stevens & Zvoch, 2006) consideran que en estos modelos los sujetos ejercen su propio control, y, por lo tanto, no requieren de variables contextuales. Pero existen opiniones contrarias que creen que la introducción de estos predictores contextuales es importante (Bryk & Raudenbush, 2002; Ferrão, 2009).

La introducción o no de predictores de contextos es un tema discutido, por un lado, hay autores que consideran que no es necesario la introducción de estas covariables porque los estudiantes ejercen su propio control (Stevens & Zvoch, 2006) o porque cuando los modelos de crecimiento se basan en el rendimiento individual del estudiante, las características de los mismos se mantienen constantes a lo largo del tiempo y no se pueden confundir con el crecimiento, y, por otro lado, autores como Raudenbush y Bryk (2002) que consideran que si es importante el uso de variables contextuales por dos motivos: el primero de ellos, porque el alumno no se ha asignado de manera aleatoria a las aulas o centros y esto puede afectar a las estimaciones y, en segundo lugar, es que si las covariables introducidas guardan

relación con el constructo a estudiar, al controlar dichos efectos se incrementará la precisión de las estimaciones.

En el caso de los modelos contextualizados un tema importante es que variables incluir en el modelo, tema complejo, pues la introducción de covariables que correlacionan con los efectos de las escuelas, pueden producir errores sistemáticos en la estimación (McCaffrey, et al., 2004). Esto se debe a que al introducir covariables, el modelo atribuye el verdadero efecto de las covariables y una parte del efecto del centro que correlaciona con dicha covariable, al efecto estimado por la covariable, y de manera equívoca elimina del efecto de los centros, la parte que correlaciona con la covariable. De esta manera, el efecto que se estima para la escuela tan sólo recibe la parte residual del efecto verdadero que no se correlaciona con la covariable. Por ejemplo, si se utiliza un alto nivel socioeconómico para ajustar un modelo donde la escuela cuenta con un VA verdadero muy alto, los efectos de la eficacia se estiman por debajo de lo que realmente son. Si se excluye la covariable del modelo se producirá un efecto opuesto. Por lo tanto, controlar las covariables de los alumnos no es suficiente para eliminar el efecto de las características contextuales.

Según la OECD (2008) el uso de variables socioeconómicas en los MVA puede tener un impacto negativo en la equidad y eficiencia de la toma de decisiones, sin embargo, también consideran que esto depende del uso que se haga con la información de las estimaciones del VA, ya que esta información puede ser muy valiosa si se utiliza, por ejemplo, para identificar a centros con bajo nivel de logro y realizar un estudio en mayor profundidad sobre los mismos con el fin de mejorar el rendimiento de los estudiantes.

En cualquier caso, parece que los factores contextuales afectan más a los modelos de ganancia que a los modelos de crecimiento. Tekwe et al. (2004) realizaron un estudio en el cual comparan cuatro modelos de ganancia, donde tres no se encontraban contextualizados y uno era contextualizado. Al analizar las correlaciones entre las diferentes estimaciones del VA en los diversos modelos, no hay muchas diferencias entre aquellos modelos que no usan variables de contexto, con valores por encima de 0,95. En el caso del modelo contextualizado, producen estimaciones diferentes con valores de correlación entre 0,50 y 0,90 dependiendo

de la asignatura y del nivel educativo evaluado. Dichos autores concluyen, que si los resultados del VA tienen como objetivo proporcionar recursos a los centros con niveles de logro bajo, no deberían utilizarse modelos contextualizados, ya que distorsionan los resultados de los centros educativos que tienen una mayor proporción de alumnado con bajo nivel socioeconómico. Sin embargo, si el fin es premiar a las escuelas con altos niveles de logro, si no se usan modelos contextualizados, estos centros saldrían perjudicados.

Hibpshman (2004) considera que los MVA que carecen de covariables de los estudiantes o de las escuelas, producen estimaciones que son fiables para identificar a escuelas o profesores que se encuentran en los extremos de la distribución, pero, sin embargo, pueden producir sesgo cuando se trabaja con poblaciones de alumnos que cuentan con altos niveles de logro.

Dentro de las covariables de contexto que se suelen utilizar, parece que el rendimiento previo y el nivel socioeconómico juegan un papel muy importante. Choi et al. (2006) estudiaron el efecto de ambas variables, donde el nivel socioeconómico se midió mediante la ayuda de comedor escolar. Estos autores, analizaron el efecto de dichas variables sobre los ranking de las escuelas. Para ello probaron un modelo con solo la variable rendimiento previo, otro con solo la variable nivel socioeconómico y un modelo que incluyó ambas variables. Al estudiar los ranking encontraron correlaciones altas entre el modelo que incluyó solo el rendimiento previo y el que incluyó el nivel socioeconómico, con valores de 0,97. Asimismo, encontraron una correlacion perfecta entre el modelo que incluyó ambas variables y el que incluyó solamente el rendimiento previo, por lo que, el estatus inicial del estudiante es capaz de captar muchos de los efectos que mide el nivel socioeconómico, lo que supone, que la utilizacion de este único predictor es suficiente para recoger una gran variedad de efectos.

Además del nivel de rendimiento previo o nivel socioeconómico de los estudiantes, el sexo y la raza también son variables de uso común en los modelos contextualizados de valor añadido, ya que permiten modelar la heterogeneidad de los resultados (Lockwood et al., 2007; Sanders & Horn, 1994). Del mismo modo, la introduccion de variables de contexto de los centros en este tipo de diseños, es muy común (Thum, 2003).

Por lo comentado, en este trabajo se considera que el uso de covariables en los MVA pueden producir sesgos en los resultados, si dichos predictores correlacionan con el efecto de la escuela. Esta cantidad de sesgo, depende de diversos factores, como la intensidad de la correlacion entre los efectos y las covariables y entre las covariables y las puntuaciones de rendimiento (McCaffrey, et al., 2003b).

La inclusión de predictores debe realizarse con el fin de reducir el sesgo de las estimaciones de los niveles de logro de las escuelas, pero hay que ser conscientes de que a veces se pueden producir los efectos contrarios. Por ello, es conveniente el estudio de la relación entre las covariables y los efectos aleatorios estimados. Si la correlación es significativa, sería adecuado la eliminación de dicho predictor, pero si no es significativa, la inclusión en el modelo no producirá sesgo. Otra posible solución, es utilizar un modelo de efectos fijos para incluir las covariables de los estudiantes y usar posteriormente las puntuaciones ajustadas, con el fin de elaborar un modelo que contenga efectos aleatorios (Ballou et al., 2004).

La introducción de variables contextuales pueden dar sentido a la estimacion de los resultados, pero se debe ser cauteloso a la hora de considerar la relación de causa-efecto entre los resultados obtenidos en los MVA y dichas variables. La relación de causalidad es un tema ampliamente discutido, sobre el que no hay acuerdo unánime acerca de si las estimaciones del VA pueden tener un valor causal o simplemente descriptivo (Ballou et al., 2004; Briggs, 2008; Kane & Staiger, 2008; Reardon & Raudenbush, 2008;Rothstein, 2009).

La estimación que los efectos de un determinado docente o un programa educativo tienen sobre los resultados de aprendizaje, y la utilización de los mismos para la toma de decisiones sobre los agentes evaluados, con el fin de aplicar sanciones o premios, supone dotar de sentido causal a los resultados cuando generalmente las relaciones de causa y efecto se encuentran vinculadas con estudios experimentales, donde los sujetos están distribuidos de manera aleatoria en los diferentes tratamientos. En el contexto educativo, los tratamientos son las diferentes escuelas, docentes, programas, etc., siendo poco probable la distribución aleatoria de los estudiantes a los mismos.

En el caso de evaluar la efectividad de los docentes, los MVA que se usan están diseñados con el fin de medir la contribución de un profesor al logro del estudiante. La eficacia se evalúa al comparar típicamente los resultados de un profesor individual, con la media de maestros eficaces en su distrito. En pocas palabras, la eficacia docente se calcula como la diferencia entre el rendimiento de los estudiantes después de estar en la clase de un determinado maestro, en comparación con lo que su logro habría sido si hubieran estado con el maestro "promedio".

Pero hay otros factores en la evaluación de los programas o de los profesores que podrían influir en el rendimiento del estudiante. Estos factores incluyen características de los estudiantes, el clima escolar, las políticas del distrito e incluso los maestros anteriores del estudiante (McCaffrey et al., 2003), todo ello son variables que pueden afectar y que hacen que sea difícil hablar de relación causal.

Además, las escuelas tienden a tener poblaciones de estudiantes con diferentes capacidades y contextos socioeconómicos, lo que dificultad aún más saber si los efectos estimados en los MVA se deben a los centros o bien a factores de contexto. Esto supone, por un lado, que es poco preciso hablar de causalidad, y, por otro, que la comparación de escuelas en términos de puntuaciones o ganancias medias, pueda ser equívoco si no se trabaja con diseños aleatorizados.

Así, si existen otros factores ajenos pero vinculados con la escuela que están afectando a los resultados académicos, estos niveles de logro no pueden atribuirse directamente a las escuelas, debido a que la ausencia de asignación aleatoria hace que muchas variables no se hayan controlado. Por ejemplo, un centro que no tenga estudiantes asignados al azar y que solo cuente con alumnos con alto poder adquisitivo, sus buenos resultados se pueden deber a dicho factor y no solo a la calidad de la enseñanza.

En determinadas ocasiones para suplir esta falta de aleatorización, se realizan ajustes mediante la introducción de variables tales como el sexo, edad, nivel socioeconómico, etc., de esta manera se equilibran las diferencias iniciales entre los estudiantes. También se pueden introducir variables de centro como titularidad, porcentaje de alumnado inmigrante, etc.; para equilibrar las diferencias a nivel de escuela. El fin de la introducción de dichos predictores es evitar el efecto de posibles variables que pueden afectar en las estimaciones de los resultados, pero, aun así,

esto no es suficiente para hablar de causalidad (Wiley, 2006). Para hablar de relaciones de causa-efecto es necesario que se den una serie de supuestos que solo se asumen en los estudios experimentales puros (Tabla II.2)

Tabla II.2. *Supuestos de los diseños causales en los MVA*

Supuesto	Descripción
Asignación aleatoria de los estudiantes a las escuelas	Los estudiantes que son asignados a los distintos docentes o escuelas, deben tener la misma probabilidad de asistir a cada una de ellas, de tal manera, que cada escuela tenga una mezcla de alumnos con similares características (Martineau, 2009; Reardon & Raudenbush, 2008)
Valor estable en la unidad de tratamiento	Todos los alumnos que se asignan a una escuela tienen que recibir los mismos estímulos, por lo que los resultados de los estudiantes son independientes entre sí (Reardon & Raudenbush, 2008)
Importancia de los datos perdidos	En determinados MVA de carácter longitudinal, los datos perdidos pueden afectar a la estimación del VA. Estos datos suelen proceder generalmente de estudiantes que cambian de centro o bien que no asisten el día de las aplicaciones.
Escala de intervalo en la distribución de las puntuaciones que provienen de los tests	La comparación causal de los centros se realiza estudiando las medias globales de diferentes unidades de análisis y buscando diferencias en los resultados. La distribuciones que provienen de los tests están en escala de intervalo (Martineau, 2009; McCaffrey et al., 2003; Reardon & Raudenbush, 2008)
Control de variables de contexto que influyen en las estimaciones	Los análisis tienen que incluir todos los factores que pueden afectar al constructo evaluado, con el fin de poder aislar el efecto de las escuelas o docentes. Sin embargo, es difícil saber que todos los factores que pueden afectar a las estimaciones han sido incluidos (Martineau, 2009)

Fuente: elaboración propia a partir de Navarro (2013)

Debido al incumplimiento de dichos supuestos, los MVA no deberían ser vistos como métodos de estimaciones causales, sino, más bien, como un método que proporciona resultados de tipo descriptivos.

El considerar estos resultados de manera causal, puede ser un problema especialmente en las evaluaciones de alto impacto (*high stakes*), ya que se pueden tomar una serie de decisiones de carácter sancionador, a partir de resultados en los cuales los efectos de ciertas variables no se han tenido en cuenta.

Rothstein (2009) mantiene que no es adecuado interpretar las estimaciones del VA de manera causal debido a la falta de aleatorización. Aunque dicho autor considera que es posible paliar dicho sesgo si se incluyen elementos de control en la asignación de los estudiantes a las aulas, teniendo en cuenta dos aspectos: el primero de ellos, si es posible observar los factores que determinan la asignación, como, por ejemplo, el rendimiento previo, y el segundo, si estas variables correlacionan con los términos del error y pueden variar con el tiempo.

Koedel y Betts (2009) estudian el problema de falta de aleatorización en los modelos de crecimiento y descubren que el sesgo de estimación se reduce en las estimaciones del VA de algunos modelos. Según dichos autores, los modelos de efectos fijos para los estudiantes en los que se evalúan el efecto de los profesores durante tres años consecutivos, reduce el sesgo que se produce por la asignación en las estimaciones de los efectos de los docentes.

Kane y Staiger (2002) estudian el efecto de la introducción de una serie de covariables en estudios no experimentales, con el fin de comparar sus resultados con un estudio experimental, donde se asigna a los docentes en los diferentes grupos de estudiantes de manera aleatoria. Estos autores concluyen que los efectos de los modelos en los que se controla las puntuaciones previas y las características de los alumnos, son un buen predictor de los resultados, ya que explican la mitad de la variación de los efectos de los profesores en los experimentos, pero se desvanecen en las siguientes evaluaciones.

Así pues, a pesar de que se pretende dotar a las estimaciones de VA de un carácter causal, esto solo se puede realizar con diseños experimentales puros, donde exista asignación aleatoria. Aunque los MVA intentan paliar esa falta de asignación mediante la introducción de diversos predictores contextuales, sería muy osado hablar de causalidad. Las estimaciones del VA añadido se deben interpretar de manera descriptiva, de tal manera que sus resultados se utilicen como indicadores para identificar escuelas que requieren de un estudio en profundidad.

CAPITULO III: ESTIMACIÓN DEL VALOR AÑADIDO EN LAS GRANDES EVALUACIONES EDUCATIVAS

El diseño de muchos estudios en las ciencias sociales requiere de individuos que se encuentran agrupados en unidades superiores (por ejemplo, escuelas, familias, comunidades, etc.). Para hacer frente a la dependencia o correlación entre los individuos de un mismo grupo, es una práctica común el uso de una metodología multinivel dentro de la cual se engloban los modelos jerárquicos lineales y de efectos mixtos (Hox, 2002; Raudenbush & Bryk, 2002; Snijders & Bosker, 1999).

La principal diferencia entre ambas metodologías reside en los parámetros que estiman, mientras que en los modelos completamente anidados (modelos jerárquicos) se estima un estatus (puntuación promedia en la primera aplicación) y una tasa de crecimiento (pendiente), en los modelos lineales mixtos se estima las ganancias que se producen de una medición a otra en términos promedio.

A pesar de las potencialidades de ambos modelos, no son válidos cuando los individuos cambian de grupo con el tiempo, por ejemplo, cuando se desea estudiar los niveles de logro de un estudiante a lo largo de varios cursos, es probable que los alumnos cambien de docente incluso de centro educativo. Dado que un modelo multinivel asume que los individuos permanecen en los mismos grupos en las diferentes ocasiones de medida, sería inapropiado su uso. Por ello, se han desarrollado modelos de efectos aleatorios clasificados de forma cruzada que admiten que la pertenencia a un grupo sea un efecto aleatorio variable en el tiempo (Goldstein, 1987; Raudenbush, 1993), es decir, estos modelos permiten estudiar la contribución que los distintos docentes realizan sobre el nivel de rendimiento de sus alumnos.

Los modelos de efectos cruzados son de suma importancia, ya que la mala especificación de un modelo que cuenta con una estructura de miembros clasificada de forma cruzada, puede conducir a estimaciones inexactas de los efectos aleatorios y de los errores estándar de los efectos fijos (Chung & Beretvas, 2012; Luo & Kwok, 2009, 2012).

Cuando no se especifica de manera correcta el modelo, el error estándar de la interacción con el tiempo de docencia es sobreestimado debido a la modelación

inadecuada de los efectos del grupo que varían en el tiempo (Luo & Kwok, 2012), lo que puede conllevar a errores estándar que son artificialmente altos, a la disminución de la potencia para detectar efectos, incluso a intervalos de confianza que son demasiado amplios. Por ello, cuando se quiere conocer cuánto de la variabilidad en el resultado es atribuible al docente, si se subestima la varianza del profesor porque se ha especificado erróneamente el modelo con respecto a la pertenencia a un grupo (Luo & Kwok, 2012), se corre el riesgo de subestimar el efecto del maestro e identificar incorrectamente a los docentes con buenos o malos resultados. Por lo tanto, es importante que el modelo se especifique con la mayor exactitud posible con respecto a la pertenencia a un grupo.

Además, estos efectos de los docentes se pueden mantener constantes a lo largo del tiempo, dando lugar a modelos acumulativos como es el caso de la *Evaluation Value Added Assessment System* (modelo EVAAS) o bien se pueden suavizar con el paso del tiempo como ocurre en el modelo de permanencia de McCaffrey, Koretz, Louis y Hamilton (2004).

Partiendo de dichas ideas, en este capítulo III se estudian los modelos de efectos anidados y de efectos cruzados, se describe el modelo acumulativo EVAAS y el de persistencia de McCaffrey et al. (2004) y, además, se analizan los modelos de efectos fijos y de efectos aleatorios en las escuelas, así como las características principales del modelado jerárquico lineal y del modelo lineal mixto.

III.1. Modelos de efectos anidados y de efectos cruzados

Generalmente, la estructura de los datos procedentes del campo educativo suele contar con una agrupación jerárquica o anidada. Pero esta no es la única estructura posible e idónea, ya que, si se pretende seguir la trayectoria de los estudiantes que cambian de aula o de centro, y, por lo tanto, de docente, se requiere el uso de modelos cruzados puesto que permiten estudiar el efecto de los distintos profesores sobre los niveles de logro de los estudiantes.

III.1.1. Efectos anidados

En un diseño longitudinal, una estructura jerárquica se caracteriza por tener una serie de observaciones que forman el primer nivel o micro nivel, que a su vez se encuentran anidadas dentro de otro nivel superior, que al mismo tiempo puede estar dentro de otro nivel o niveles superiores.

En el campo educativo, y, concretamente, dentro de los modelos de valor añadido, generalmente las ocasiones de medidas son consideradas el nivel inferior y se encuentran anidadas en los estudiantes, que a su vez se anidan en las aulas, tal y como se representa en la figura siguiente:

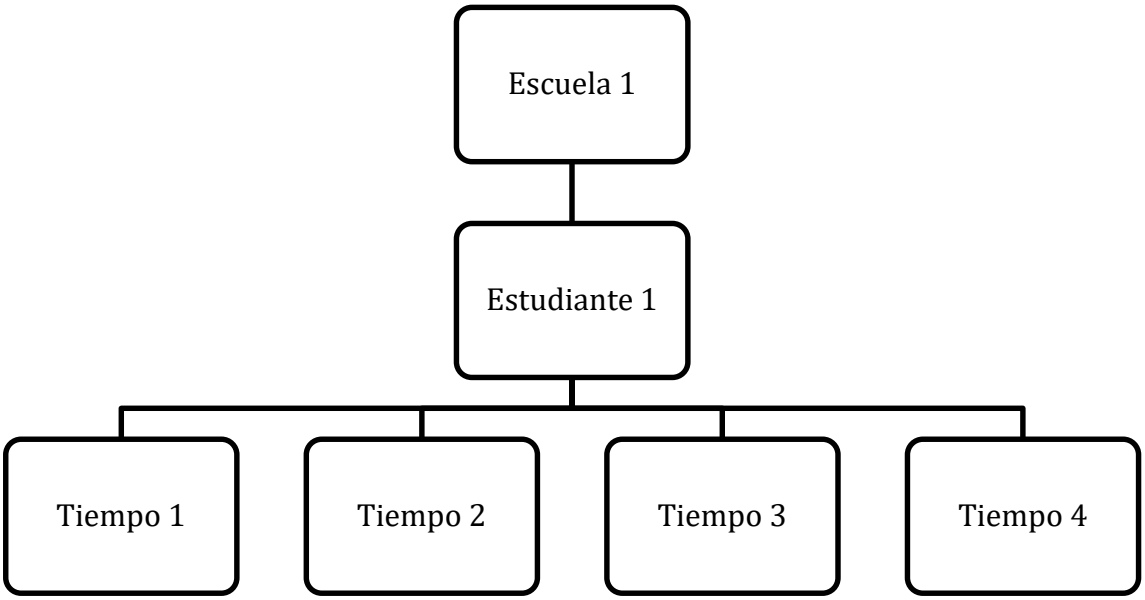


Figura III.1. *Ejemplo de estructura jerárquica*
Fuente: elaboración propia

En la Figura III.1, se puede observar que todas las ocasiones de medida (tiempo 1, tiempo 2, tiempo 3 y tiempo 4), se encuentran anidadas en el estudiante 1, que a su vez se encuentra anidado en la escuela a la que pertenece (escuela 1).

Como se ha estudiado en el capítulo II, los datos del sector educativo se caracterízan generalmente por tener una estructura anidada, que viola el supuesto de independencia requerido por los análisis tradicionales tales como el ANOVA, análisis de regresión múltiple, etc. Esto se debe a que, por ejemplo, las puntuaciones de una variable respuesta, como puede ser el nivel de logro de los estudiantes que

pertenecen a la misma escuela, es probable que correlacionen más que las calificaciones de los estudiantes que asisten a diferentes centros educativos, ya que comparten el mismo entorno. Estas violaciones independentistas hacen que los modelos multinivel se conviertan en una necesidad, ya que los análisis tradicionales pueden producir excesivos errores de tipo I y estimaciones sesgadas de los parámetros (ver Aitken & Longford, 1986; Brown, 1994; Burstein, 1980; Cronbach, 1976; Cronbach & Webb, 1975; Goldstein, 1988; Raudenbush & Willms, 1995)

En este sentido, los modelos multinivel o modelos jerárquicos lineales son considerados como la forma más adecuada para calcular las medidas de VA. Con ellos, la naturaleza anidada de los datos educativos se respeta, reconociendo explícitamente diferentes niveles y fuentes de variabilidad (Martínez-Arias, 2009).

Con estos modelos es posible distinguir qué parte del logro se debe a cada nivel, proporcionando estimaciones más fiables del efecto de las escuelas (OECD 2008). Otra ventaja de los métodos multinivel, es que, aunque la mayoría de las veces proporcionan un modelado lineal de crecimiento, cuando se consideran más de tres puntos temporales, una relación no lineal puede representar mejor la evolución del aprendizaje, algo que estos métodos permiten modelar (CCSSO 2008; Reckase, 2008).

III.1.2. Efectos cruzados

Si los estudiantes cambian de escuela o de docente, es necesario utilizar un modelo multinivel de efectos cruzados (Choi et al., 2006; Choi & Seltzer, 2005; Goldstein, 1987; 2003; Goldstein, Burgess & McConnell, 2007; Ponisciak & Bryk, 2005; Raudenbush & Bryk, 2002).

Cuando se tiene en cuenta los cambios de centro o de maestro que se producen en los estudiantes, dichos cambios se pueden considerar como efectos cruzados. Esto es lo más correcto, ya que dentro del sistema educativo es normal que los alumnos tengan distintos docentes incluso en un mismo curso. El problema es que este tipo de análisis complica en exceso el diseño, pues, a no ser que se utilicen modelos muy específicos como es el caso del modelo EVAAS, se suele estimar un único residuo asociado al crecimiento de los docentes, cuando estos atienden a alumnos diferentes en cada medida.

Lo más interesante de estos modelos cruzados, es que permiten estimar la contribución de cada docente sobre los niveles de logro de los estudiantes, algo que si no se tiene en cuenta puede dar lugar a estimaciones finales del VA sesgadas (Daniel, 2012; Grenn, 2010; Meyers & Beretvas, 2006).

Existen varios modelos para estimar los efectos de los maestros, que capturan estas complejidades de los resultados de los estudiantes. Rowan, Correnti y Miller (2002) realizaron una comparación empírica de los efectos estimados de los maestros, en los resultados longitudinales obtenidos en los exámenes de una muestra de alumnos. Para ello utilizaron tres modelos distintos: modelos de ajuste de covariable con las puntuaciones actuales regresadas a las puntuaciones anteriores; ganancias de un año (es decir, puntuación del año actual menos puntuación del año anterior) y un complejo modelo de efectos cruzados con efectos aleatorios. Los autores informan que, a través de múltiples sujetos y cohortes, la contribución de los maestros a la variabilidad total en las puntuaciones, varió de un 4 a un 16 por ciento para los modelos de ajuste de covariable, de 3 a un 10 por ciento para los modelos de puntuación de ganancia y de 10 a un 20 por ciento para los modelos de efectos cruzados. Rowan et al. (2002) argumentan a favor del modelo cruzado porque, a diferencia de los otros modelos, descomponen la variabilidad del crecimiento dentro del aula en sus componentes sistemáticos y residuales.

Estos autores proporcionan una primera exploración importante de la robustez de los efectos del maestro, aunque su estudio es limitado, ya que no realizan un análisis detallado sobre los supuestos del modelo o sobre los posibles resultados en el caso de la violación de estos supuestos. Además, no investigan si los MVA realmente eliminan los efectos de factores, tales como el desempeño previo y el estatus socioeconómico y proporcionan un indicador más preciso de la efectividad del maestro de lo que es posible, cuando estos factores no están controlados.

Los modelos de efectos cruzados, a pesar de su complejidad, parecen adecuados, pues cuando se trabaja con datos longitudinales los resultados estudiantiles obtenidos en los MVA presentan muchos desafíos a la hora de modelar los efectos y la variabilidad entre los docentes. Las múltiples medidas realizadas a un mismo estudiante que cuentan con la influencia de distintos profesores tiene que

ser contemplado, ya que es probable la aparición de una correlación positiva entre las medidas múltiples en el mismo estudiante.

Los grupos de estudiantes cambian anualmente de clase y los estudiantes son enseñados por un profesor diferente cada año. La existencia de efectos del profesor y variables ambientales, compartidas, pero no observadas para los estudiantes dentro de las aulas, contribuirá a una correlación intraclase positiva en los resultados de los estudiantes de la misma clase.

Estos cambios de centro o de maestros a lo largo de los años, significa que los resultados de los estudiantes no siguen los diseños anidados tradicionales (Goldstein, 2003; Raudenbush & Bryk, 2002) pues en dichos diseños jerarquizados cuando un sujeto cambia de centro se elimina de la evaluación, mientras que en los modelos de efectos cruzados al contemplar los cambios de escuela o de profesores no cuentan con tantos valores perdidos (Hill & Goldstein, 1998).

En la siguiente tabla se representa la estructura de datos completamente anidados.

Tabla III.1. *Estructura anidada de los datos longitudinales*

Alumno/Escuela	A	B	C	D
1	XXXX			
2		XXX		
3			XXX	
4			XX	
5				XXXX

Fuente: elaboración a partir de Navarro (2013)

En la tabla anterior cada fila es un estudiante y cada columna una escuela, representándose un total de cinco estudiantes distribuidos en cuatro escuelas diferentes. Las X se refieren a las mediciones realizadas, siendo el número máximo de ocasiones de medida cuatro. Cada una de ellas se encuentran anidadas en un estudiante, que a su vez se anida en una escuela. Cada ocasión se realiza en la misma

escuela, ya que no se contempla la posibilidad de que los estudiantes cambien de centro, por lo tanto, el sujeto 1 pertenece a la escuela A, el 2 a la B, el 3 y 4 a la C y el 5 a la D.

Sin embargo, cuando se habla de modelos de efectos cruzados si es posible que los alumnos cambien de escuela o de profesores a lo largo de la evaluación, tal y como se puede observar en la Tabla III.2

Tabla III.2. *Estructura cruzada de los datos longitudinales*

Alumno/Escuela	A	B	C	D
1	XX	X	X	
2		X	X	X
3			XX	XX
4			X	X
5	X		XXX	

Fuente: elaboración a partir de Navarro (2013)

En este caso, las mediciones se anidan en los estudiantes, pero los sujetos y las escuelas se encuentran en un único nivel de manera cruzada, y no en dos como en el ejemplo anterior. Estos modelos si tienen en cuenta la posibilidad de que un alumno pueda cambiar de aula y, además, son capaces de captar ese seguimiento, pues, por ejemplo, el alumno 1 es evaluado dos veces en la escuela A, uno en la B y otra en la C.

Las aplicaciones de VA basándose en los modelos de efectos cruzados, debido a sus ventajas han atraído el interés de un gran número de investigadores y analistas, que han encontrado en estos modelos una posible aplicación sobre las evaluaciones individuales de los maestros. Algo muy relevante para el estudio del VA, ya que los docentes tienen efectos grandes y diferenciales en el aprendizaje de los estudiantes.

Entre los distintos modelos basados en el estudio de los efectos cruzados, cabe destacar el *Sistema de Evaluación de Valor Agregado de Tennessee* (TVAAS), denominado posteriormente como *Evaluation Value Added Assessment System* (EVAAS), desarrollado por Sanders en 1993. Dicho método permite seguir a los estudiantes cuando cambian de maestro, ya sea en distintos cursos o dentro de un mismo curso.

Para el análisis de dichos modelos, al no tratarse de modelos anidados no utilizan los análisis jerárquicos lineales, sino que se basan en los modelos lineales mixtos[9] con efectos cruzados, que a diferencia de los jerárquicos tienen mayor dificultad para estimar el efecto de los factores de los centros educativos y comunidades, sobre el crecimiento de los estudiantes a lo largo del tiempo; ya que se centran principalmente en el estudio de los efectos cruzados de los docentes, las aulas o las escuelas. Añadir covariables para el estudio de estos factores sería algo muy costoso en términos de cálculo y estimación.

Desde la perspectiva de efectos cruzados, en los modelos jerárquicos lineales también es posible estimar el efecto que distintos docentes o escuelas tienen sobre los niveles de logro de los estudiantes. A estos modelos se les denomina como modelos de clasificación cruzada (*cross classified*) *(*Bryk & Raudenbush, 2002).

En los modelos *cross classified,* el crecimiento se analiza de la misma manera que en los modelos jerárquicos lineales utilizados en la estimación del VA. Así cada medición se estimaría tal y como aparece en la siguiente ecuación (Navarro, 2013):

Ec.III.1

$$Y_{1i} = \beta_0 + T_0(\beta_1 + r_{1i}) + r_{0i} + u_{1j} + e_{1i}$$

$$Y_{2i} = \beta_0 + T_1(\beta_1 + r_{1i}) + r_{0i} + u_{1j} + u_{2j} + e_{2ij}$$

$$Y_{3i} = \beta_0 + T_2(\beta_1 + r_{1i}) + r_{0i} + u_{1j} + u_{2j} + u_{3j} + e_{3ij}$$

En la ecuación anterior Y_{ti} hace referencia a la puntuación en la ocasión *t* para el estudiante *i*, pero contemplando la posibilidad de que en cada momento temporal los centros educativos pueden tener distintos estudiantes. β_0 es la media global del

[9] Estos modelos serán estudiados con mayor profundidad en el apartado III.4.2

intercepto, es decir, la media global en la primera ocasión de medida y β_1 es la tasa de crecimiento (pendiente) vinculada al término lineal de crecimiento (T), que en este ejemplo al considerar tres ocasiones de medida toma los valores 0, 1 y 2 (Navarro, 2013).

La ecuación del modelo general queda planteada de la siguiente manera (Bryk & Raudenbush, 2002):

$$Y_{tij} = \beta_{00} + r_{0i} + T_{ti}(\beta_{10} + r_{1i}) + \sum_{j=1}^{J}\sum_{t=0}^{T} Du_{t(ij)} + e_{tij} \qquad \textbf{Ec.III.2}$$

En esta ecuación los coeficientes aleatorios son los *r* y el $u_{t(ij)}$. El coeficiente r_{0i} hace referencia al residuo asociado al intercepto y el r_{1i} a la pendiente. Ambos se distribuyen de forma normal con matriz de varianza-covarianza.

$$\begin{pmatrix} r_{0i} \\ r_{1i} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{r0} & \\ \sigma_{r1,r0} & \sigma^2_{r1} \end{pmatrix}\right] \qquad \textbf{Ec.III.3}$$

El otro efecto aleatorio, el $u_{t(ij)}$ es el de la escuela y esta cruzado con los estudiantes, ya que refleja la posibilidad de que los mismos cambien de maestros entre las aplicaciones. Dicho residuo son las desviaciones esperadas en el crecimiento, que tienen lugar cuando el estudiante se encuentra con el profesor *j* en la ocasión de medida *t*. Este efecto es acumulativo, y, por lo tanto, se va sumando a lo largo de las mediciones en un mismo estudiante, así, la puntuación del alumno se debe tanto al efecto del docente actual como de los previos (Bryk & Raudenbush, 2002). $u_{t(ij)}$ también cuenta con una distribución normal con media cero y matriz de varianzas-covarianzas constante entre aplicaciones.

La variable D se refiere a la acumulación de los efectos de los docentes, y toma el valor de 1 cuando el alumno recibe la enseñanza del profesor *j* en la aplicación *t*, y el valor cero en caso contrario (Bryk & Raudenbush, 2002).

Finalmente el término e_{tij} se refiere al error intra-sujeto que cuenta con media normal y varianza constante a lo largo del tiempo $N(0, \sigma^2)$.

III.2. Efectos acumulativos y de persistencia suavizada

El modelo EVAAS es un modelo de efectos cruzados que permite estudiar la influencia de los distintos docentes manteniéndose dicho efecto constante, aun cuando el alumno cambia de maestro. De esta manera los niveles de logro de los estudiantes, dependen de los efectos de los maestros actuales, pero también de los que han tenido en las mediciones anteriores.

Este modelo parte de la premisa de que el aprendizaje de los estudiantes se acumula, siendo los docentes quienes provocan los incrementos en los niveles de logro de sus pupilos, por ello, la influencia de todos los docentes debe ser considerada como permanente.

Esta característica de permanencia es lo que hace que el modelo EVAAS sea también conocido como de "*layered model*" o modelo estratificado o por capas, ya que los efectos de los docentes anteriores influyen en los resultados actuales. Por lo tanto, la varianza de las puntuaciones, debido a la contribución de los maestros va aumentando con el tiempo, aunque la varianza total no se vea incrementada (McCaffrey et al., 2004)

Otros modelos afines se basan en una suavización de los efectos de los docentes de los años anteriores, incluso contemplan la posibilidad de que estos efectos se eliminen con el tiempo. Así, permiten una disminución de los efectos previos de los docentes en las estimaciones del año estudiado. La metodología de estos modelos es similar a la del modelo de Sanders (1991) solo que incluye un parámetro que se estima en cada momento, y permite la variación de los efectos aleatorios previos de los maestros o los centros educativos. Este modelo es conocido como modelo de persistencia y fue desarrollado por McCaffrey et al. (2003b).

En definitiva, en estos diseños en los que un estudiante cambia de docente en cada aplicación pueden darse varias posibilidades que se van a estudiar a continuación: modelos donde los efectos de los docentes se acumulan, como es el caso del modelo de Tennesse y modelos donde el efecto de los docentes se suaviza o incluso desaparece de un año a otro.

III.2.1. El modelo de valor añadido de Tennessee (EVAAS)

El Sistema de Evaluación de Valor Añadido de Tennessee (EVAAS) mide el impacto que las escuelas y los profesores tienen en el progreso académico de sus estudiantes. EVAAS se centra en el estudio del crecimiento estudiantil, no en si el estudiante es competente en la evaluación del estado. De esta manera, dicha evaluación ayuda a los educadores a identificar las mejores prácticas y a implementar los programas que mejor satisfacen las necesidades de sus alumnos, así como, a tomar decisiones informadas sobre dónde enfocar los recursos, con el fin de asegurar oportunidades de crecimiento para todos los estudiantes.

Una parte fundamental de esta evaluación es una base de datos masiva fusionada longitudinalmente, que vincula los estudiantes y los resultados de los estudiantes con las escuelas y los sistemas en los que están inscritos, y con los maestros que se les asignan a medida que pasan de grado a grado.

Las investigaciones realizadas utilizando la base de datos del EVAAS han demostrado que la efectividad del maestro, es el principal determinante del progreso académico del estudiante. Han comprobado que los efectos de los maestros sobre el rendimiento estudiantil son tanto aditivos como acumulativos, con poca evidencia de si los maestros eficaces pueden compensar los efectos de los docentes ineficaces (Sanders & Horn, 1998).

En el modelo EVAAS las puntuaciones de los sujetos se estiman de la siguiente manera (Navarro, 2013):

Ec.III.4

$$Y_{1i} = \beta_1 + Pu_{1j} + r_{1i}$$

$$Y_{2i} = \beta_2 + Pu_{1j} + Pu_{2j} + r_{2i}$$

$$Y_{3i} = \beta_3 + Pu_{1j} + Pu_{2j} + Pu_{3j} + u_{3i}$$

$$Y_{4i} = \beta_4 + Pu_{1j} + Pu_{2j} + Pu_{3j} + Pu_{4j} + r_{4i}$$

En la ecuación anterior, Y_{ti} hace referencia a la puntuación en el test en un determinado momento *t* (generalmente un año específico) del estudiante *i*; β_t hace referencia a la media global de ese momento *t*; u_{tj} es el efecto del docente *j* en el

momento *t*, es decir, el valor añadido del maestro. Dicho VA son las desviaciones respecto a la media global, que se distribuye de manera normal, con media cero y varianza común en cada momento; P es la medida de proporción de escolarización de un alumno con un determinado maestro, si ese profesor no le ha impartido docencia adquiere el valor de 0; r_t hace referencia al residuo en el año *t*, se distribuye de manera normal con una matriz de varianzas-covarianzas intra-estudiantes, que permite diferentes patrones de correlacionen entre las aplicaciones pero no entre los sujetos (Navarro, 2013).

Tal y como señala el autor anterior, este modelo incorpora un coeficiente fijo para cada momento de medida, así como efectos aleatorios cruzados, utilizando un complejo sistema de matrices que permite contemplar el cambio de los alumnos entre los diferentes docentes, incluyendo efectos previos de resultados actuales.

EVAAS emplea tres análisis separados de modelos lineales mixtos: para los distritos, las escuelas y los profesores. Los modelos de distrito y escuela son esencialmente los mismos.

En el caso de los docentes el modelo permite estudiar el efecto de muchos maestros, con relativamente pocos alumnos por maestro. El modelo queda definido de la siguiente manera (Navarro, 2013):

$$Y_{tij} = \beta_t + \sum_{t=1}^{T} Pu_{t(ij)} + r_{tij} \qquad \textbf{Ec.III.5}$$

Mediante la ecuación anterior III.5, se estima las medias globales (parámetros fijos) en las cuatro ocasiones de medida que se realizan y como componente aleatorio el residuo asociado a cada estudiante. Los efectos cruzados de los diferentes maestros se establecen mediante la estimación de un residuo, para cada uno de ellos en cada ocasión de medida, utilizando una matriz de coeficientes que permite el cambio entre los maestros en un mismo curso y entre los cursos. Asimismo, los errores residuales de los alumnos (r_{tij}) que se encuentran asociados a cada puntuación de rendimiento, pueden estar correlacionados para un mismo alumno y de esta manera captar la estructura longitudinal de la variable dependiente.

En la siguiente Tabla III.3, se muestra un ejemplo de los coeficientes aleatorios utilizados para estimar los efectos de los docentes. Se trata de un ejemplo de tres estudiantes que tienen tres escenarios diferentes de maestros durante los grados 3º, 4º y 5º, en tres materias: Matemática (M), Lectura (R) y Lenguaje (L). Los maestros son identificados por las letras de la A a la F.

Tabla III.3. *Ejemplo de matriz de coeficientes*

			PROFESOR																	
			SEGUNDO GRADO						TERCER GRADO						CUARTO GRADO					
			A			B			C			D			E			F		
GRADO	ASIGNATURA	ESTUDIANTE	M	R	L	M	R	L	M	R	L	M	R	L	M	R	L	M	R	L
2º	M	TOMMY	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	R		0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	L		0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3º	M		1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
	R		0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
	L		0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
4º	M		1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
	R		0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0
	L		0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0
2º	M	SUSAN	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	R		0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	L		0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3º	M		1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	R		0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
	L		0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
4º	M		1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0
	R		0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0
2º	M	ERIC	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
	R		0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
	L		0	0	.5	0	0	.5	0	0	0	0	0	1	0	0	0	0	0	0
3º	M		1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
	R		0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0
	L		0	0	.5	0	0	.5	0	0	0	0	0	1	0	0	0	.2	0	0
4º	M		1	0	0	0	0	0	0	0	0	1	0	0	.8	0	0	0	0	0
	R		0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0
	L		0	0	.5	0	0	.5	0	0	0	0	0	1	0	0	.5	0	0	.5

Fuente: elaboración propia a partir de Wright, White, Sanders & Rivers (2010)

Como se observa en la tabla anterior, los maestros de Tommy representan el escenario convencional: Tommy es enseñado por un solo maestro en las tres asignaturas cada año (maestros A, C y E en grados de segundo, tercero y cuarto, respectivamente). Nótese que, en la matriz de Tommy, para el tercer curso hay unas filas que representan la presencia de un efecto de profesor, no sólo para el maestro de tercer grado C, sino también para el maestro de segundo grado A (así, es cómo se codifica la "estratificación"). De manera similar, en las filas del cuarto grado se encuentra la influencia del maestro E, el maestro C en el grado tercero y el maestro A en el segundo grado.

Susan es enseñada por dos profesores diferentes en el grado 2º, el maestro A para las Matemáticas y el Lenguaje y el maestro B para la Lectura. En el grado 3º, Susan recibió clase del maestro C, para la Lectura y el Lenguaje. Por alguna razón, en el grado 3º ningún maestro enseñó a Susan Matemáticas, a pesar de que Susan tenía una puntuación en la prueba de Matemáticas del tercer grado. Esta puntuación, todavía se puede incluir en el análisis mediante la introducción de ceros en las filas de la matriz de Susan, para Matemáticas del 3er grado. Además, en el grado 4º, Susan no tenía puntuaciones en el Lenguaje, esta fila se omite completamente de la matriz debido a la falta de puntuaciones.

El escenario de Eric ilustra la enseñanza en equipo. En la asignatura de Lenguaje de segundo grado, Eric recibió la misma cantidad de instrucción de dos profesores (maestros A y B). Las entradas en la matriz indican la contribución de cada uno de ellos, 0,5 para cada maestro. En este caso, Eric recibe la docencia de dos docentes distintos, y esos efectos permanecen constantes en las otras aplicaciones. Del mismo modo, en la asignatura de Lenguaje de cuarto grado, Eric recibió clase de dos profesores por igual, maestros E y F. En el cuarto grado de Matemáticas, aunque Eric fue enseñado por ambos profesores E y F, no hicieron una contribución igual. El maestro E tenía un 80% de responsabilidad y el maestro F tenía el 20%.

En estos modelos se pretende analizar las diferencias entre dos grados adyacentes e identificar el efecto de cada docente, asumiendo que el efecto previo del docente se mantiene a lo largo del tiempo. De esta manera, el efecto del profesor hace referencia a la ganancia que queda una vez que se elimina la ganancia del estudiante, la ganancia en la media global y la contribución de posibles factores relativos a las características del estudiante si se incluyeran (Ballou et al., 2004):

En estos modelos lineales mixtos, los efectos de los maestros son tratados como efectos aleatorios mediante técnicas de estimación conocidas como estimaciones bayesianas o BLUP. Esto significa, que a priori un maestro se considera "promedio" (con un efecto de cero) hasta que haya suficientes datos de los estudiantes como para indicar lo contrario. Este método de estimación, protege contra falsos positivos (maestros incorrectamente evaluados como efectivos) y falsos negativos (maestros incorrectamente evaluados como ineficaces), particularmente en el caso de maestros con pocos alumnos (Wright et al., 2010)

El principal problema de este tipo de estimaciones, tal y como se estudiará en el siguiente capítulo, es que cuando se utilizan para estimar el efecto de las escuelas, se basan en los datos de otras escuelas para estimar el efecto que tiene un determinado centro educativo, produciendo el conocido efecto (*shrunk o shrunkage*)

III.2.2. Modelo de persistencia

McCaffrey et al. (2003), sugieren una alternativa al modelo por capas (EVAAS), que permite estimar la fuerza de la persistencia de los efectos del profesor en los últimos años. El modelo se encuentra definido por la siguiente ecuación:

Ec.III.6

$$\mathrm{Y}_1 = \beta_1 + T_1 + e_{i1}$$

$$\mathrm{Y}_2 = \beta_2 + T_2 + a_{21}T_1 + \mathrm{e}_{\mathrm{i}2}$$

$$\mathrm{Y}_3 = \beta_3 + T_3 + a_{32}T_2 + a_{31}T_1 + \mathrm{e}_{\mathrm{i}3}$$

Este modelo es conocido como modelo de persistencia, porque modela la persistencia de los efectos del profesor estimando los parámetros, a_{21}, a_{31} y a_{32} que determinan la persistencia de los efectos de los maestros del año anterior, en las puntuaciones del año en curso. Así, mientras que en el modelo EVAAS los efectos de los docentes se consideran constantes a lo largo del tiempo, en el de persistencia los efectos de los maestros se vuelven a estimar en los cursos siguientes, quedando dicho efecto suavizado.

Ambos modelos tienen en común que permiten que las varianzas de los errores residuales, cambien a lo largo de los años y las correlaciones entre los términos de errores no se especifican y se estiman a partir de los datos.

En el modelo de persistencia, la ganancia que se produce de un año a otro, tiene en cuenta los efectos de los docentes previos de manera suavizada, dicha ganancia se estima mediante la siguiente ecuación (McCaffrey et al., 2003b):

$$\mathrm{Y}_3 - \mathrm{Y}_2 = (\beta_3 - \beta_2) + T_3 + (a_{32} - 1)T_2 + (a_{31} - a_{21})T_1 + e_3 - e_2 \qquad \textbf{Ec.III.7}$$

En la ecuación anterior las ganancias del segundo grado al tercer grado, dependen del maestro de segundo grado a través del término $(a_{32} - 1)T_2$ y del maestro de primer grado a través del término $(a_{31} - a_{21})T_1$. Con el modelo de persistencia de McCaffrey et al. (2003) los efectos de los maestros de los grados 1 y 2 afectan tanto a las puntuaciones del grado 3º, como a las ganancias del grado 2º al grado 3º. En particular, cuando a $a_{32} < 1$, si T_2 es positivo, el maestro de segundo grado tiene una contribución negativa sobre las ganancias, y viceversa si T_2 es negativo. Esta es una forma de regresión a la media porque los efectos del maestro se "desgastan" con el tiempo y, las puntuaciones de los estudiantes se movilizan hacia el desempeño que ya no esta afectado por el maestro de segundo grado. Por el contrario, cuando a $a_{32} > 1$, si T_2 es positivo, entonces el maestro de segundo grado tiene un efecto positivo en el crecimiento durante el tercer grado y viceversa si T_2 es negativo. Cuando a_{32} es mayor que uno, los efectos previos de los maestros aceleran el crecimiento y los maestros ineficaces desaceleran el crecimiento. Resultados similares se aplican a los otros parámetros a (McCaffrey et al., 2003b)

Finalmente, hay que señalar que dicho modelo utilizado para los profesores podría ser utilizado con otros efectos de la escolarización. Sin embargo, hasta el momento la mayoría de los modelos longitudinales han ignorado los efectos escolares del año anterior en los modelos de las puntuaciones del año en curso, suponiendo implícitamente que los efectos decaen a cero en un año. Los modelos de ajuste de covarianza suponen que persisten sólo cuando contribuyen a las puntuaciones de los exámenes del año anterior, es decir, sus efectos se descomponen al mismo ritmo que los efectos de todos los otros atributos en el logro previo. No hay exploraciones empíricas de la robustez de las estimaciones a las suposiciones sobre los efectos de la escolaridad de los años anteriores.

Debido a que es probable que el tratamiento de los efectos escolares tenga una gran influencia en los efectos estimados del maestro, en algunas poblaciones (ver McCaffrey et al., 2003b), se espera que las estimaciones sean sensibles al modelado de los efectos de la escolaridad del año anterior.

Los autores del modelo de persistencia, también desarrollaron un modelo longitudinal para las puntuaciones de los estudiantes, que incluye los efectos del docente, centro y posibles covariables. El modelo se formula teniendo en cuenta que

los datos de la puntuación son, Y_{ig} para la puntuación del estudiante *i* en el grado *g*. Para comodidad de notación, se considera *g* = 0 para el primer grado de recolección de datos, *g* = 1 para el segundo y así sucesivamente. El modelo para los puntajes del grado 0 quedaria definido de la siguiente manera (McCaffrey, et al., 2004):

$$Y_{i0} = \mu_0 + \beta_0 x_i + \gamma_{00} z_{i0} + \sum_{K=1}^{M} \lambda_{i0k}\eta_{0k} + \sum_{j_0=1}^{N_0} \phi_{i0j}\theta_{0j} + \epsilon_{i0} \qquad \textbf{Ec.III.8}$$

El η_{0k} denotan los efectos escolares del grado 0 (es decir, las desviaciones en las medias del nivel escolar de la media general del sistema) aunque también puede ser con respecto a algún estándar establecido previamente. El λ_{i0k} mide la proporción de escolaridad del grado 0, que la escuela proporcionó al estudiante *i*. Si el estudiante no asistió a la escuela *k*, entonces λ_{i0k}= 0. Si el estudiante asistió sólo a la escuela *k* entonces $\lambda_{i0k} = 1$, de lo contrario, está entre 0 y 1.

Para los estudiantes que pasan sólo parte del año en el sistema educativo, la suma de $\lambda_{\lambda i0k}$ será menor que 1. Para los estudiantes que completan todo el año en este sistema, la suma de λ_{i0k} será igual a 1. Del mismo modo, los θ_{0j} se refieren a los efectos de los maestros y ϕ_{i0j} miden la proporción de educación del grado 0, proporcionada al estudiante *i* por el maestro *j*. Los valores de λ_{i0k} y ϕ_{i0j} se observan a partir de datos administrativos, y no se calculan.

El μ_0 es la media de todos los estudiantes en ese grado, y el x_i y z_{i0} son covariables fijas y covariables que pueden variar en el tiempo, respectivamente. Por ejemplo, género, raza (fijas) o circunstancias especiales durante la prueba (variantes). Además, es posible la inclusión de variables referidas a la escuela como la titularidad, porcentaje de alumnado inmigrante, etc. Así, β y γ son los coeficientes de regresión asociados a los predictores.

Este modelo aumenta aún más su complejidad según se va incluyendo factores que afectan a las puntuaciones de rendimiento o variables dependientes de los niveles de logro.

El modelo estadístico presentado anteriormente, caracteriza los efectos del docente como variables aleatorias que contribuyen a las puntuaciones de los exámenes. Sin embargo, los efectos de interés del profesor, son contribuciones

causales de los maestros al logro estudiantil. La relación entre el modelo estadístico y el efecto causal depende de numerosos supuestos.

En este modelo se encuentra implícito que el profesor tiene un efecto constante en todos los estudiantes, en relación con otros maestros en el sistema. Dado que los efectos del docente pueden no ser constantes, el efecto es una aproximación al efecto promedio del profesor sobre los estudiantes de la población que probablemente estén en su clase, suponiendo que el modelo está correctamente especificado.

En el primer año de pruebas, es poco probable que el modelo se especifique correctamente porque no tiene en cuenta la historia del estudiante antes del grado en el que se realiza la prueba. En este caso, el efecto estimado del profesor debería tender a incluir factores históricos tales como: antecedentes estudiantiles y experiencia educativa previa, que se agrupan por aula. Así, los efectos estimados del profesor cuando $g = 0$, deben ser interpretados con cautela. Esto también ocurre para los efectos escolares.

Además, en este modelo se consideran dos medidas de los efectos del profesor: estimaciones de los efectos individuales de los maestros y la contribución general de los maestros a la variabilidad en los resultados de los estudiantes. Para la estimación de los efectos de los docentes, este modelo también se basa en las estimaciones BLUP, ya que proporcionan, dentro de los modelos mixtos, estimaciones no sesgadas del efecto de cada maestro individual (Searle, Casella & McCulloch, 1992).

En resumen, estos métodos de modelado del VA proporcionan oportunidades para estimar la proporción de variabilidad en el logro o el crecimiento estudiantil atribuible a los maestros, así como estimar el efecto de un maestro individual sobre el aprendizaje del estudiante. Cuando estos métodos identifican grandes diferencias en la efectividad de los maestros, también tienen el potencial de ayudar a los investigadores a identificar qué características poseen los docentes altamente efectivos y motivar a mejores prácticas (McCaffrey et al., 2003b).

Las estimaciones del efecto del profesor pueden usarse para una gran variedad de propósitos, desde informar a los sistemas educativos acerca de cómo los estudiantes se ven afectados por las prácticas y condiciones actuales, hasta

tomar decisiones de alto riesgo en relación con el sueldo y / o el empleo del maestro. Sin embargo, aunque los métodos de modelado con valor añadido infieren en los efectos causales de los profesores sobre el crecimiento estudiantil, los datos de la evaluación no se obtienen de estudios experimentales aleatorios. En consecuencia, existen varias limitaciones al concretar lo que los efectos del profesor realmente describen.

Definir los efectos de los maestros, requiere identificar el impacto de un profesor particular sobre el crecimiento del estudiante en el logro, que posteriormente se comparará con otros docentes en la escuela, distrito o estado. La definición también depende de los resultados utilizados para medir el logro. El alcance y el propósito de los instrumentos puede limitar lo que se mide y, en consecuencia, restringir la parte del impacto total de un profesor sobre las estimaciones del estudiante (McCaffrey et al., 2003b). Otros factores que afectan al crecimiento de los estudiantes en el rendimiento, tales como las características de las aulas y las escuelas, pueden confundirse con las estimaciones de los efectos del profesor, por lo que, el propósito de obtener estas estimaciones debe estar claramente definido y deben determinar con qué precisión los efectos deben ser estimados. Típicamente, los efectos del profesor simplemente explican la heterogeneidad inexplicada en el aula (Lockwood et al., 2007).

Los estudios que investigan los efectos del maestro en los modelos de valor añadido, proporcionan pruebas de que los docentes tienen diferentes efectos sobre el aprendizaje de los estudiantes que persisten a lo largo del tiempo (Rowan et al., 2002; Wright, Horn & Sanders, 1997) pero estos estudios tienen defectos (McCaffrey et al., 2003b) y requieren de análisis más profundos.

III.3. Efectos fijos y aleatorios

Dentro del marco de los MVA utilizados en las evaluaciones educativas, en este apartado se estudiará el tratamiento de los efectos de las escuelas cuando son considerados de manera fija o aleatoria.

Acorde con la literatura basada en los diseños experimentales, los efectos fijos se refieren a aquellos en los que las inferencias solo se pueden realizar sobre los términos específicos que se incluyen en el modelo. Sin embargo, se habla de

efectos aleatorios cuando los elementos seleccionados se consideran una muestra aleatoria de la población, e interesa analizar la variabilidad existente en la población de aquellas unidades que han sido muestreadas (Martínez-Arias et al., 2009). En términos generales, el uso de una aproximación u otra depende de las inferencias que se deseen extraer de los datos, pero un enfoque no es considerado mejor que otro.

Las aplicaciones tempranas de los MVA (por ejemplo, Hanushek, (1972) y Murnane, (1975)) utilizaron principalmente modelos basados en los efectos fijos, mientras que aplicaciones más recientes, han utilizado modelos de efectos aleatorios. Generalmente, para estimar el VA se utiliza un modelado mixto donde los efectos de la escuela son considerados como aleatorios (Ballou et al., 2004; Goldstein, 2003; McCaffrey et al., 2004; Sanders, et al., 1997) siendo estos de gran utilidad cuando se trabaja con estructuras de datos que se correlacionan con la heterogeneidad del intercepto y el cambio en los centros educativos (Martínez-Arias et al., 2009).

Al tratar los efectos de las escuelas como variables aleatorias, los resultados referentes a la variabilidad de las estimaciones, son similares a los obtenidos cuando se consideran de efectos fijos. Pero el uso diferenciado de efectos aleatorios y efectos fijos, puede afectar a los resultados de las estimaciones finales del VA, ya que tratan de manera diferente al error muestral (Martínez-Arias, et al., 2009; Sanders & Wright, 2008; Tekwe et al., 2004). Los efectos fijos se encuentran extremadamente afectados por el error muestral, cuando se trabaja con aulas o escuelas que cuentan con tamaños muestrales muy pequeños.

En los modelos de efectos fijos el efecto de la escuela se estima a partir de los alumnos del centro, mientras queen los modelos de efectos aleatorios se usan las estimaciones bayesianas, conocidas como estimaciones BLUP que contraen (*shrunk*) el estimador que se basa en los estudiantes, hacia la media global del mismo. Estos estimadores sesgados, aunque no para todos los autores (por ejemplo, Wright at al., 2010) poseen propiedades psicométricas adecuadas, menos para las escuelas cuyos efectos se encuentran lejos de la media (Martínez-Arias et al., 2009).

III.3.1. Modelos de efectos fijos de las escuelas

Cuando los parámetros son tratados como efectos fijos, significa que las unidades que se observan son las únicas que interesan y, por lo tanto, son consideradas como la población (McCaffrey et al., 2003b).

En los MVA cuando se considera a los centros educativos como variables de efectos fijos, las escuelas que se incluyen en la muestra son toda la población de interés y, por lo tanto, no existe varianza aleatoria entre las mismas.

El modelo de efectos fijos más sencillo es el modelo de regresión simple, en el cual se asume que los coeficientes fijos son comunes a todos los sujetos y se estiman a partir de los datos.

En la siguiente ecuación se muestra un modelo de regresión simple en el cual no se tiene en cuenta el crecimiento en los niveles de logro.

Ec.III.9

$$Y_i = \beta_0 + e_i$$

$$e_i \sim N(0, \sigma_e^2)$$

En la ecuación anterior, Y_i se refiere al nivel de rendimiento del estudiante *i*, el coeficiente β_0 es el promedio de toda la población y el e_i es la diferencia entre la puntuación predicha de un estudiante y su puntuación observada, es decir, el residuo. Los residuos se consideran que son independientes, con media cero y varianza común para todos los estudiantes.

Dicho residuo puede expresarse de la siguiente manera:

$$e_i = Y_i - \widehat{Y}_i$$ **Ec.III.10**

Donde Y_i es la puntuación observada del estudiante *i* y $\widehat{Y}_i$ su puntuación predicha en la recta de regresión. La diferencia entre la puntuación observada y la predicha es el error.

La estimación del valor predicho $\widehat{Y}_i$ del estudiante *i,* de una escuela *j*, se realiza con todos los datos disponibles de la muestra. Por lo tanto, si se quiere estimar el valor añadido se calcularía como la media de los errores de los estudiantes de un centro educativo. Los estudiantes de una escuela concreta, que obtienen niveles de

logro altos en las puntuaciones finales (en comparación con alumnos de otros centros con características afines), sus residuos serán positivos y producirán una estimación del VA añadido positivo para esa escuela.

III.3.2. Modelos de efectos aleatorios de las escuelas

Los coeficientes aleatorios se distribuyen acorde a una función de probabilidad. En un modelo multinivel con dos niveles de anidamiento (por ejemplo, estudiantes y centros), los coeficientes del nivel 1 (alumnos) son tratados como coeficientes aleatorios en el nivel de escuela (nivel 2). Dichos modelos permiten a los grupos desviarse de la solución central tanto en el estatus como en la pendiente.

Cuando se asumen modelos de efectos aleatorios, las unidades de interés son una muestra aleatoria de una población similar pero no observable, por lo tanto, la variabilidad entre los elementos observados describe la varianza de la población.

En los modelos en los que se considera la contribución de las escuelas como efectos aleatorios, se encuentran determinados por la propia escuela, pero también por los datos de todos los otros centros.

El modelo más básico de efectos aleatorios, se encuentra formado por dos ecuaciones de regresión, referidas, en este caso, al nivel de estudiantes y al nivel de centro. Siendo la ecuación de regresión del primer nivel los estudiantes y del segundo nivel las escuelas.

Las ecuaciones de este modelo serian la siguiente:

Nivel 1 (estudiantes) $$Y_{ij} = \beta_{0j} + e_{ij}$$ **Ec.III.11**

Nivel 2 (escuelas) $$\beta_{0j} = \beta_{00} + r_{0j}$$ **Ec.III.12**

El nivel 1 hace referencia al alumno *i*, del centro *j*. El coeficiente β_{0j}, es fijo y se refiere a la media de los estudiantes de la escuela *j* y finalmente el e_{ij} es el efecto aleatorio, ya que alude al residuo de cada estudiante.

El coeficiente fijo del nivel de estudiante, pasa a formar parte de la ecuación referida al nivel de escuela (segundo nivel). En esta ecuación, el β_{00} se refiere a la media global en los niveles de logro de todas las escuelas y r_{0j} es el residuo, es decir, la desviación de un centro concreto (*j*) respecto a la media global.

Las desviaciones o residuos en ambas ecuaciones son consideradas independientes con distribución normal, tal y como se muestra a continuación.

$$e_{ij} \sim N(0, \sigma_e^2)$$
$$r_{0j} \sim N(0, \sigma_r^2)$$

Ec.III.13

Normalmente, los MVA se estiman mediante estos modelos mixtos o lo que es lo mismo, modelos multinivel que están compuestos por dos partes. Una parte más general referida a la parte común de los contextos, conocida como parte fija y otra parte que representa lo específico de cada contexto. Esta parte, es la parte aleatoria, puede variar y se estima a través de la varianza de los distintos niveles (Murillo, 2008). La metodología más utilizada para estimar estos modelos mixtos dentro de los MVA, es como es de esperar, los modelos lineales mixtos y/o modelos jerárquicos lineales que forman parte de estos primeros.

III.4. Modelos jerárquicos lineales y modelos lineales mixtos

El VA no se puede comprender sin las técnicas de análisis que se emplean para estimarlo.

Goldstein (1991) dijo que la única forma de comparar escuelas es estudiando el progreso de los estudiantes durante su permanencia en el centro educativo. El progreso del estudiante se puede medir al comparar las diferencias de año a año, por ello, la metodología más adecuada para medir los cambios en el rendimiento de los estudiantes, es a través de la estimación de las trayectorias de crecimiento individual, mediante un modelo multinivel (Bryk & Raudenbush, 1987; Willett, 1988; Willms, 1992).

En este sentido, las técnicas para la estimación del cambio son variadas, aunque las más comunes son los modelos jerárquicos lineales (Singer & Willett, 2003) existen también otras posibilidades como, los modelos lineales mixtos o las ecuaciones estructurales. En este trabajo se explicarán los Modelos Jerárquicos Lineales porque es la metodología más utilizada en las evaluaciones a gran escala y los Modelos Lineales Mixtos con medidas repetidas, ya que es la técnica empleada en la conocida evaluación EVAAS.

Dichas técnicas multinivel permiten la conceptualización de las medidas de resultados más relevantes, es decir, el estudio de la eficacia de la escuela, al facilitar la estimación del beneficio adicional o "valor", que los estudiantes recibirán al asistir a un colegio en particular (Boyle & Willms, 2001; Bryk & Raudenbush, 1988; Willms, 1992).

El modelado multinivel facilita las estimaciones del rendimiento escolar en términos de valor añadido al proporcionar un cierto grado de control sobre una gran cantidad de factores de confusión.

Todo esto, junto con la naturaleza anidada de los datos educativos ha dado lugar a que la modelización multinivel se aplique cada vez más en este campo, ya que dichas técnicas evitan problemas como el sesgo de agregación y la sobreestimación de los errores estándar (Raudenbush & Bryk 2002).

Los modelos multinivel o jerárquicos (Raudenbush & Bryk 2002) al tener en cuenta la naturaleza anidada de los datos dentro de unidades más grandes, formulan ecuaciones para cada nivel de agregación, pudiendo mostrar la relación entre las variables dentro de un determinado nivel. Así, es posible distinguir el logro debido a las diferentes fuentes de variación, proporcionando estimaciones más precisas de la incertidumbre que se adjunta a las estimaciones del VA de las escuelas (OECD, 2008). De esta manera, estos métodos permiten comprobar la adecuación de los modelos matemáticos, que permiten resumir las relaciones entre las variables medidas dentro de las diferentes agrupaciones, que forman una estructura de datos jerárquica.

Uno de los beneficios de los modelos multinivel, es que se incluyen las puntuaciones de todos los estudiantes para estimar el crecimiento, tanto de los individuos como los grupos, incluso si el estudiante tiene datos incompletos. A excepción de los modelos de ganancia, que con sólo dos puntuaciones se requieren de datos completos.

Otra fortaleza de este enfoque, es que debido a su sofisticada técnica estadística empleada, puede maximizar el uso de la información disponible y minimizar algunos tipos de error al proporcionar una estimación precisa del rendimiento de los estudiantes.

A la hora de estimar los efectos de las escuelas se pueden utilizar diferentes modelos estadísticos. Lo importante es que se incluyan al menos tres o más mediciones, sobre los niveles de logro de los estudiantes, pudiéndose incluir o no, variables contextuales de los estudiantes y de los centros educativos. Los distintos modelos pueden tener supuestos distintos, y utilizar diferentes algoritmos a la hora de realizar las estimaciones, por lo que los resultados pueden diferir en cierta medida.

Mediante los modelos jerárquicos lineales se estima el estatus inicial o intercepto (rendimiento promedio en la primera medición) y una pendiente de crecimiento a lo largo del tiempo (Raudenbush & Bryk 2002) con sus residuos correspondientes. Mientras que, a través de los modelos lineales mixtos con medidas repetidas, se estiman tantos coeficientes y residuos como aplicaciones de medida (Sanders & Horn, 1994).

Aunque ambos modelos difieren en los parámetros que estiman, a través de ellos se pretende estudiar el crecimiento de los estudiantes en términos de niveles de logro a lo largo del tiempo (Wiley, 2006)

Una ventaja tanto de los modelos jerárquicos lineales, como de los modelos lineales mixtos con medidas repetidas, es que, se puede especificar distintas estructuras de matrices de varianza-covarianza, que permiten detallar las características de los datos en los distintos niveles (Garson, 2013; Pardo & Ruiz, 2013; Raudenbush & Bryk, 2002; Ronald, Scott & Lynn, 2010; Rosário, et al. 2011).

La especificación correcta de la hipótesis del tipo de covarianza utilizada para generar los modelos lineales mixtos con medidas repetidas, y los modelos jerárquicos lineales, es importante porque el modelado mixto utiliza un algoritmo iterativo para estimar los coeficientes, que hace que las estimaciones sean más fiables si el algoritmo utiliza como punto de partida una suposición precisa, acerca de la naturaleza de la matriz de varianzas y covarianzas de las variables en el modelo.

Suponiendo una estructura de covarianza demasiado sencilla aumentará los errores de Tipo I, al tiempo que asumir una estructura demasiada compleja aumentará los errores Tipo II, cuando se interpreten los componentes de la varianza para los efectos aleatorios en la salida de los modelos (Garson, 2013).

Es importante tener en cuenta la existencia de dos tipos de dominios para los supuestos de estructura de covarianza. En el primer dominio de efectos aleatorios, referido generalmente al nivel de alumno (cuando se trabaja con dos niveles) se especifican los efectos aleatorios de la matriz de varianza-covarianza "Matriz G"; a partir de la cual se estiman los efectos inter-sujetos. El segundo dominio es para medidas repetidas, que especifica la estructura de varianza- covarianza de esas medidas, "Matriz R". Esta es la base para la estimación de los efectos intra-sujetos. La estructura de covarianza especificada para los efectos aleatorios (Matriz G) no tiene porque ser la misma que la establecida para medidas repetidas (Matriz R).

De tal manera que puede ocurrir que no exista un mismo tipo de matriz de covarianza especificada para cada nivel (Garson, 2013). En SPSS, R y otros programas, es posible ejecutar modelos multinivel donde cada nivel de anidamiento cuente con su propia estructura de covarianza especificada.

Dependiendo del tipo de datos, sobretodo del número de mediciones y la distancia existente entre las mismas, la selección de un tipo u otro de matriz puede dar lugar a poca o ninguna diferencia en los resultados. Sin embargo, cuando se trabaja con pocas mediciones en el tiempo, la mala especificación del tipo de estructura de covarianza puede conducir a un error sustancial en las conclusiones del investigador (Garson, 2013).

A continuación, se presenta una tabla con las matrices de varianza-covarianza más utilizadas en los modelos multinivel.

Tabla III.4. *Estructura de matrices de varianza-covarianza en los modelos multinivel*

Matriz	Estructura	Características
Sin estructura	$\begin{pmatrix} \sigma_1^2 & & \\ \sigma_{21} & \sigma_2^2 & \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{pmatrix}$	- Cada elemento en la matriz puede tener un valor diferente - Esta opción es adecuada cuando el investigador no tiene ninguna base para saber cuál es la estructura de covarianza - El problema de dicha estructura, es que el modelo que genera es el menos parsimonioso de todos por la gran cantidad de parámetros que tiene que estimar
Diagonal	$\begin{pmatrix} \sigma_1^2 & & \\ 0 & \sigma_2^2 & \\ 0 & 0 & \sigma_3^2 \end{pmatrix}$	-Tiene varianzas heterogéneas y una correlación de cero entre cada ocasión de medida, por lo que asume que no hay covarianza entre las medidas, siendo las mismas independientes
Identidad	$\begin{pmatrix} 1 & & \\ 0 & 1 & \\ 0 & 0 & 1 \end{pmatrix}$	- Es la matriz que utilizan ciertos programas por defecto si no se especifica otra - Se trata de una matriz diagonal que tiene una varianza constante y asume que no existe correlación alguna entre los elementos. Por lo tanto, contiene en la diagonal principal el mismo valor y ceros fuera de la misma (Pardo & Ruiz, 2013)
Simetría Compuesta	$\begin{pmatrix} \sigma^2+\sigma_1^2 & & \\ \sigma_1 & \sigma^2+\sigma_1^2 & \\ \sigma_1 & \sigma_1 & \sigma^2+\sigma_1^2 \end{pmatrix}$	- Asume que las varianzas de las medidas repetidas son iguales entre sí (diagonal principal), e iguales entre sí también, las covarianzas entre cada par de medidas (valores iguales de la diagonal principal, pero no necesariamente ceros) - Es la más parsimoniosa de todas (incluye el menor número posible de parámetros), pero no siempre es la mejor elección cuando se trabaja con medidas repetidas (Pardo & Ruiz, 2013). - En este tipo de matriz, la correlación de los residuos para las mediciones cercanas en el tiempo, son las mismas que para las mediciones que están muy alejadas - Derivado de la matriz anterior, se encuentran: la **matriz de simetría compuesta: métrica de correlación**, que tiene varianzas homogéneas y correlaciones homogéneas entre los elementos y la **matriz de simetría compuesta: heterogénea,** que asume varianzas heterogéneas y una correlación constante entre los elementos

Autorregresiva (AR1)	$\begin{pmatrix} \mathbf{1} & & \\ \boldsymbol{\rho} & \mathbf{1} & \\ \boldsymbol{\rho}^{\mathbf{2}} & \boldsymbol{\rho} & \mathbf{1} \end{pmatrix}$	- Permite modelar una estructura donde las medidas más próximas correlacionan más entre sí que las más alejadas. Al considerarse esta estructura como la esperable en los modelos con medidas repetidas, es la que se ha venido utilizando tradicionalmente en los diseños longitudinales - Cuenta con varianzas homogéneas en la diagonal y con un parámetro de correlación (ρ) para las covarianzas. Este parámetro disminuye su valor a medida que se alejan las aplicaciones. Por tanto, es igual a ρ en el caso de elementos adyacentes y ρ^2 cuando se trata de elementos separados entre sí por un tercero, y así sucesivamente. ρ ésta limitado de manera que $-1 < \rho < 1$ - Otros dos tipos de matrices autorregresiva son: **autorregresiva: heterogénea**, que se trata de una estructura autorregresiva de primer orden, con varianzas heterogéneas, es decir, similar a la matriz AR1, pero asume que los elementos diagonales pueden ser diferentes y **ARMA** que se trata de una estructura de media móvil autorregresiva, con varianzas homogéneas - La estructura AR1 es un caso particular de una estructura más general llamada **Toeplitz**. En esta matriz el valor de las covarianzas depende de la distancia entre las medidas repetidas, pero no se obtienen a partir de un único coeficiente de correlación intraclase, como en AR1, sino que se estima un coeficiente de correlación para cada distancia

Fuente: elaboración propia a partir de Ronald et al. (2010)

Estos tipos de matrices pueden ser utilizadas, tanto en los modelos jerárquicos lineales como, en los modelos lineales mixtos. Además, ambas metodologías cuentan con una serie de características comunes, por ejemplo:

- Son modelos cuantitativos que utilizan las puntuaciones de los estudiantes en los tests, como medida de los niveles de logro.
- Requieren de más de dos mediciones para estimar el logro académico, con la finalidad de conocer cuánto cambio se produce de una aplicación a otra (diseños longitudinales).
- En ambas metodologías, es necesario que las puntuaciones se encuentren en una escala común para que se puedan comparar a lo largo del tiempo. Generalmente, para ello, se utilizan escalas verticales de rendimiento.
- A través de dichos modelos, se pretende determinar qué cantidad de cambio producido en el nivel de rendimiento del estudiante, puede ser atribuido al centro o al maestro.
- Estos modelos tienen la posibilidad de incorporar, o no, variables de contexto de los estudiantes o del centro.

Asimismo, también se diferencian en varios aspectos (Rogosa & Willet, 1983):

- Mientras que en los modelos jerárquicos lineales, se representan las trayectorias del crecimiento individual, mediante la estimación de un estatus (puntuación promedia en la primera aplicación) y una tasa de crecimiento (pendiente), más el error aleatorio con media cero y varianza común entre aplicaciones. En los modelos lineales mixtos, se estima la puntuación promedia en cada una de las aplicaciones, donde los residuos correspondientes a cada una de las ocasiones pueden estar correlacionados.
- Otra distinción, es que los modelos lineales mixtos con medidas repetidas requieren que las mediciones se encuentren espaciadas de la misma manera, puesto que se basan en cálculos de ganancia. Sin embargo, los modelos completamente anidados, son más laxos respecto a las características de los datos, ya que al considerar las medidas repetidas dentro de un individuo pueden variar, tanto en número, como distancia temporal.
- Finalmente, en los modelos lineales mixtos, añadir covariables supone mucho coste en términos de procesos de cálculo y estimación, mientras que

en los modelos jerárquicos se pueden estudiar factores del centro o de los estudiantes a través del tiempo incluyendo un tercer nivel.

III.4.1. Modelos jerárquicos lineales completamente anidados

Los modelos multinivel (Goldstein & deStavola, 2010; Hox, 2010; Snijders & Bosker, 2012, Garson, 2013), conocidos también como modelos jerárquicos lineales completamente anidados (Bryk & Raudenbush, 1992; Raudenbush & Bryk, 2002), modelos de efectos o coeficientes aleatorios (De Leeuw & Kreft, 1986; Longford, 1993) o modelos de componentes de la varianza (Longford, 1989), son una extensión de los modelos de efectos mixtos descritos por Rao (1965) para las curvas de crecimiento y por Laird y Ware (1982) para el análisis de datos longitudinales.

Los Modelos Jerárquicos Lineales (HLM, por sus siglas en inglés), son un tipo especial de modelos de la familia de los modelos mixtos generales, y pueden usarse para obtener medidas de valor añadido. Como su propio nombre indica requieren del uso de datos anidados que se ordenen jerárquicamente. Debido a la naturaleza de los datos utilizados en la educación, los modelos jerárquicos lineales han sido ampliamente adoptados por los educadores en la evaluación de las escuelas. En este campo educativo, a la hora de estimar los MVA las mediciones se encuentran anidadas en los estudiantes y los estudiantes están anidados dentro de escuelas.

Debido a la existencia de homogeneidad entre los sujetos dentro de las jerarquías, es conveniente que cada nivel cuente con su propio modelo donde están representadas las relaciones estructurales, así como la variabilidad residual (Bryk & Raudenbush, 1992). En definitiva, se trata de que el primer nivel o micro-nivel debe ajustarse para cada contexto (macro-nivel) (De la Orden, 1985). Los coeficientes de regresión del primer nivel, se analizan como variables aleatorias en el segundo nivel, descomponiéndose la varianza en la de ambos niveles.

Las diferentes puntuaciones de resultados se anidan en cada individuo, por tanto, la trayectoria de cambio se representa como una función específica de cada sujeto, más un error aleatorio (Goldstein & Woodhouse, 2001; Raudenbush & Bryk, 2002; Willett, 1988). A su vez, los estudiantes pueden estar agrupados en aulas y/o escuelas, incluso en niveles superiores.

Una fortaleza del enfoque de modelado multinivel, es que, maximiza el uso de la información disponible y minimiza algunos tipos de error, ya que es posible obtener información muy precisa sobre el rendimiento de los estudiantes. Además, los modelos de crecimiento longitudinales desde la perspectiva jerárquica, permiten examinar las trayectorias de crecimiento individual y hacer inferencias posteriores sobre la efectividad de programas que se han planteado y aplicado durante algún tiempo (Heck, 2000; Willms & Raudenbush, 1989). Su uso es cada vez mayor, gracias a su potencia de cálculo (véase, por ejemplo, Goldschmidt, 2002; Heck, 2000; Ramírez, Yuen, Ramey & Pasta, 1991) que permite la especificación de un modelo adecuado, que capta correctamente la estructura del crecimiento, dependiente de la naturaleza anidada de los datos que se está modelando (Raudenbush, 2004).

El mayor inconveniente de estos modelos, son los supuestos subyacentes que generalmente se ocultan en una "caja negra " (OECD, 2008, p.73) de complejidad que pocas personas pueden entender, y, por lo tanto, son difíciles de implementar y/o explicar. Además, otra limitación importante es el tamaño muestral, acorde con Kreft y De Leeuw (1998) se recomienda al menos 30 sujetos por aula, otros autores (Goldstein, 1997, Bryk & Raudenbush, 2002; Raudenbush & Bryk, 1986; Snijders & Bosker 2012) mencionan que los centros educativos con tamaños inferiores a 20 sujetos pueden ser un problema porque su pendiente de crecimiento tenderá a no diferenciarse de la media global. Como no existe un acuerdo unánime acerca del número de sujetos necesarios para la obtención de estimaciones precisas, en dicha tesis se estudiará cómo afecta el tamaño muestral en la estimación de la pendiente de crecimiento de las aulas. Pues, en los diseños longitudinales debido, entre otros factores, a la mortandad experimental, a veces es complejo contar con más de 20 estudiantes por aula.

Para entender mejor las características estadísticas de estos modelos multinivel, como ejemplo, se desarrolla un modelo jerárquico lineal utilizando la anotación de Raudenbush y Bryk (2002). Se describe un modelo de tres niveles (tiempo, estudiantes y escuela) en el cual se han realizado tres mediciones distintas a una misma cohorte de estudiantes. De este modo, la variable tiempo se encuentra definida como 0, 1 y 2 haciendo referencia a cada una de las aplicaciones.

El modelo queda definido de la siguiente manera:

Nivel 1 (modelo de crecimiento individual)

$$Y_{tij} = \beta_{0ij} + \beta_{1ij}(t) + e_{tij} \qquad \textbf{Ec.III.14}$$

En la ecuación anterior, Y_{1ij} se refiere al nivel de rendimiento en un momento concreto, determinado por el tiempo (t), 0, 1 y 2 del alumno *i,* en un centro *j*. El coeficiente β_{0ij}, es el intercepto o estatus inicial de partida (rendimiento promedio en la primera aplicación) del estudiante *i,* en la escuela *j* y β_{1ij} se refiere a la pendiente o tasa de crecimiento. Esta tasa de crecimiento debe multiplicarse por la variable tiempo definida previamente (t).

En dicho nivel el error residual (e_{tij}) cuenta con una distribución normal con media de cero y varianza común entre aplicaciones:

$$e_{tij} \sim N(0, \sigma^2) \qquad \textbf{Ec.III.15}$$

Nivel 2 (alumnos)

Ec.III.16

$$\beta_{0ij} = \beta_{00j} + r_{0ij}$$

$$\beta_{1ij} = \beta_{10j} + r_{1ij}$$

$$r_{ij} \sim N(0, R)$$

Ahora los coeficientes del nivel del tiempo pasan a ser una ecuación en este segundo nivel. Aquí el coeficiente β_{0j}, se refiera a la media del estatus inicial de los estudiantes en el centro *j*; β_{1j} es la media de las pendientes en esa escuela *j*. Por último, r_{0ij} y r_{1ij} son los residuos o desviaciones de los estudiantes, vinculados al estatus y a la pendiente, respectivamente. Ambos se distribuyen de forma normal con medias cero y matriz de varianzas-covarianzas R, que en este caso es una matriz sin estructura.

$$R = \begin{bmatrix} \sigma^2_{r00} & \\ \sigma^2_{r10} & \sigma^2_{r11} \end{bmatrix} \qquad \textbf{Ec.III.17}$$

Nivel 3 (escuelas):

$$\beta_{00j} = \beta_{000} + u_{00j}$$
$$\beta_{10j} = \beta_{100} + u_{10j} \qquad \textbf{Ec.III.18}$$

En este último nivel, se representan las medias globales de todo el conjunto de la muestra, tanto en el estatus inicial (β_{000}) como en la pendiente (β_{100}). Estas medias cuentan con residuos asociados a cada escuela, que informan de la aportación diferencial respecto a la media global del intercepto (u_{00j}) y la tasa de cambio de los niveles de logro que se encuentran asociados al tiempo (u_{1j}).

Si se sustituye todas las ecuaciones anteriores en una sola ecuación, el modelo quedaría definido de la siguiente manera:

$$Y_{tij} = \beta_{000} + \beta_{100}(T) + u_{00j} + u_{10j}(T) + r_{0ij} + r_{1ij}(T) + e_{tij} \qquad \textbf{Ec.III.19}$$

En la ecuación anterior se diferencian dos partes. La parte fija del modelo, compuesta por el estatus inicial (β_{000}), es decir, las medias globales en toda la muestra y la tasa de crecimiento (β_{100}) a lo largo del tiempo (T), expresado como 0 (primera medición), 1 (segunda medición), 2 (tercera medición) y así sucesivamente. La otra parte es la parte aleatoria, referida a los residuos aleatorios de cada uno de los niveles vinculados con la función de tiempo. Los coeficientes r_{0ij} $+r_{1ij}(T)$ son los residuos vinculados al nivel de estudiante y, representan las diferencias de cada estudiante respecto a la media global en estatus y pendiente. Los residuos del nivel de centro son $u_{00j} + u_{10j}(T)$, que expresan las diferencias de cada escuela respecto a la media global de las mismas en estatus y pendiente. En el nivel del tiempo se encuentra el residuo intrasujeto $e_{tij,}$ caracterizado por tener una distribución normal con media cero y varianza común entre aplicaciones.

III.4.2. Modelos lineales mixtos con medidas repetidas

Los modelos multivariados de medidas repetidas o Modelos Lineales Mixtos con medidas repetidas (MLM), tienen su caso más representativo, dentro del contexto de las grandes evaluaciones internacionales, en el modelo de EVAAS (Sanders & Horn, 1994), donde, tal y como se ha estudiado en el apartado III.2.1, se

utilizan con el fin de aislar los posibles efectos de los profesores y las escuelas, en el aprendizaje de los estudiantes.

En estos modelos lineales mixtos, a diferencia de los modelos jerárquico lineales, no se estima una pendiente de ganancia, sino que está implícita ya en el modelo al calcular una puntuación promedia en cada aplicación.

Nuevamente, a modo de ejemplo, y con fines explicativos, se plantea un modelo de crecimiento basado en los modelos lineales mixtos donde se incluyen tres puntuaciones del rendimiento como variable criterio y tres niveles de anidamiento (tiempo, estudiante y escuela).

Las ecuaciones de los distintos niveles quedan definidas de la siguiente manera:

Nivel 1: nivel de tiempo

$$Y_{1ij} = \beta_{0ij} + \beta_{1ij} + \beta_{2ij} + e_{tij} \qquad \textbf{Ec.III.20}$$

En este caso cada uno de los coeficientes β, son las medias de los estudiantes del centro *j*, en cada una de las aplicaciones (t = 0, 1 y 2) y el término $e_{tij,}$ es el residual. Este residuo de primer nivel incluye una matriz de varianzas-covarianzas (matriz E), sin estructura, que se distribuye de forma normal con media cero.

$$e_{tij} \sim N(0, E) \qquad \textbf{Ec.III.21}$$

$$E = \begin{pmatrix} \sigma^2_{e0} & & \\ \sigma_{e20} & \sigma^2_{e1} & \\ \sigma_{e30} & \sigma_{e31} & \sigma^2_{e2} \end{pmatrix}$$

El nivel 2 (estudiantes)

Ec.III.22

$$\beta_{0ij} = \beta_{00j} + r_{0ij}$$

$$\beta_{1ij} = \beta_{10j} + r_{1ij}$$

$$\beta_{2ij} = \beta_{20j} + r_{2ij}$$

$$r_{ij} \sim N(0, R)$$

En el nivel dos, no se estima una pendiente de crecimiento como tal, pero si es posible calcular las ganancias producidas entre cada una de las aplicaciones.

En la ecuación anterior III.22, los coeficientes β son las medias globales en cada una de las mediciones llevadas a cabo. Los residuos de los estudiantes asociados a cada coeficiente estimado son r_{0ij}, r_{1ij} y r_{2ij} , respectivamente. Estos errores (r_{tij}) asociados a cada puntuación de rendimiento, pueden estar correlacionados para un mismo sujeto y, de esta forma, captar la estructura longitudinal de la variable dependiente. Los residuos se distribuyen de forma normal con media cero y matriz de varianzas-covarianzas R.

$$R = \begin{pmatrix} \sigma^2_{r0} & & \\ \sigma_{r20} & \sigma^2_{r1} & \\ \sigma_{r30} & \sigma_{r31} & \sigma^2_{r2} \end{pmatrix} \qquad \textbf{Ec.III.23}$$

Nivel 3: aulas

Ec.III.24

$$\beta_{0j} = \beta_{000} + u_{00j}$$

$$\beta_{1j} = \beta_{100} + u_{10j}$$

$$\beta_{2j} = \beta_{200} + u_{20j}$$

$$r_j \sim N(0, U)$$

$$U = \begin{pmatrix} \sigma^2_{u0} & & \\ \sigma_{u20} & \sigma^2_{u1} & \\ \sigma_{u30} & \sigma_{u31} & \sigma^2_{u2} \end{pmatrix}$$

Este nivel representa las medias globales ($\beta_{000,}$ β_{100} y β_{200}) de toda la muestra, en cada una de las mediciones realizadas. Las medias tienen residuos aleatorios asociados a cada centro ($u_{0j,}$ u_{1j} y u_{2j}).

Si se sustituye las ecuaciones anteriores de los tres niveles en una única ecuación, quedaría formulada de la siguiente manera:

$$Y_{tij} = \beta_{000} + \beta_{100} + \beta_{200} + r_{0ij} + r_{1ij} + r_{2ij} + u_{0j} + u_{1j} + u_{2j} + e_{tij} \qquad \textbf{Ec.III.25}$$

Nuevamente, la parte fija del modelo ($\beta_{000} + \beta_{100} + \beta_{200}$) representan las medias globales de toda la muestra, pero en vez de hacer referencia al estatus y la pendiente, hacen referencia a la media global en cada una de las tres mediciones realizadas. En la parte aleatoria del modelo, se encuentran los residuos aleatorios asociados a las medias del nivel de estudiante ($r_{0ij} + r_{1ij} + r_{2ij}$), del nivel de escuela ($u_{00j} + u_{10j} + u_{20j}$) y el error residual del primer nivel (e_{tij}), que se distribuye de forma normal con media cero y matriz de varianza covarianza.

CAPITULO IV: LA FIABILIDAD EN LAS ESTIMACIONES DE LAS PUNTUACIONES DE LOS ESTUDIANTES EN LOS MODELOS LONGITUDINALES

A partir de la ley NCBL (2001), las evaluaciones educativas basadas en la rendición de cuentas ("*accountability systems*"), cada vez han ido adquiriendo una mayor importancia, siendo los diseños longitudinales en los que se siguen a los mismos sujetos en diferentes ocasiones de medida (modelos de ganancia o crecimiento) los que proporcionan una mayor información sobre el logro de los estudiantes. En ambos modelos la mejora del rendimiento de los estudiantes y, también, de las escuelas se juzga por los cambios que se producen en los niveles de logro de los educandos en las diferentes mediciones realizadas, pero muchas veces estos cambios se encuentran afectados por errores de medición, de muestreo, que pueden originar el conocido efecto de regresión a la media.

En este capítulo, se pretende describir el comportamiento de este fenómeno estadístico, y explicar el efecto que el mismo puede tener en la fiabilidad con la que se estiman las puntuaciones de los estudiantes, en los modelos multinivel de carácter longitudinal. Para ello, se estudia el concepto del efecto de regresión a la media y su influencia en los diseños con dos ocasiones de medida y con más de dos aplicaciones, la relación entre los patrones de correlación generados entre las puntuaciones de los estudiantes y el efecto de regresión a la media, los factores que afectan a este artefacto estadístico y las posibles estrategias para paliarlo. Finalmente, se analizan otros posibles fenómenos que, pueden afectar a la fiabilidad de las estimaciones en las puntuaciones de los estudiantes a lo largo del tiempo.

IV.1. El efecto de regresión a la media

El Efecto de Regresión a la Media (ERM, de aquí en adelante) conlleva a que los sujetos que parten de las puntuaciones más bajas tendrán tasas de cambio mayores que los sujetos con puntuaciones iniciales más altas, y, por consiguiente, los sujetos con puntuaciones más altas tendrán tasas de cambio menores que los sujetos con puntuaciones más bajas (Jamieson 1995, 1999; Lacey & Lacey, 1962; Nesselroade, et al.,1980; Raykov, 1995; Wilder, 1957)

La relación entre el estatus inicial y el crecimiento a lo largo del tiempo ha sido considerada un aspecto metodológico un tanto complejo, estudiado dentro de los modelos de ganancia (dos puntuaciones) y/o de crecimiento (más de dos puntuaciones) (Rogosa, 1995; Willett, 1997). Ya Rogosa en el año 1995 señalaba en su clásico artículo centrado en los mitos de la investigación longitudinal, que existen distintos tipos de relaciones entre el estatus y el crecimiento: positivas, negativas y nulas. A pesar de que Rogosa señala que dicho efecto puede ser un artefacto del diseño, ya que este valor de correlación está influido por la elección de un determinado punto de partida y, además, el número de ocasiones de medida pueden modificar esta relación entre el estatus y la pendiente, diversos autores (Lacey & Lacey, 1962; Nesselroade, et al.,1980; Raykov, 1995; Wilder, 1957) han considerado que la existencia de una correlación negativa entre el estatus y la pendiente, puede ser considerada como un posible indicador del ERM.

Lo cierto es que independientemente del tipo de relación existente (positiva, negativa o nula) siempre existirá un efecto de regresión a la media entre las dos primeras ocasiones de medida si no existe una correlación perfecta entre las mismas (Nesselroade et al., 1980), y como la línea de regresión entre dichas ocasiones no puede ser nunca perfecta, es inevitable encontrarse con dicho efecto (Campbell & Kenny, 1999), por lo menos en los diseños pretest-postest.

El ERM en el campo educativo ha sido estudiado sobre todo por Kahneman y Tversky (1973), Kelley (1947), Lord y Novick (1968) y Thorndike (1963), quienes observaron que la regresión a la media está muy presente en dicho ámbito, pero que en la mayoría de las ocasiones pasa desapercibido por ser considerado un fenómeno que en muchas situaciones es inevitable y que, como bien apuntan Lee y Smith (2002) una noción correcta de la regresión hacia la media es extremadamente difícil de adquirir.

En este sentido, el ERM es especialmente indeseable dentro del marco de los MVA, ya que las escuelas tienden a categorizarse de acuerdo a su valor añadido y su crecimiento. Por ello, los administradores y los responsables políticos deben trabajar para crear sistemas de rendición de cuentas, que consideren la posible existencia de ciertos artefactos estadísticos, como es el ERM, que puede enturbiar la estimación del verdadero nivel de logro de los estudiantes sino se elimina. El

problema, es que, más allá de los estudios realizados por Furby (1973) y Nesselroade et al. (1980), no existen investigaciones recientes que expliquen qué es el ERM y cómo se comporta en las evaluaciones educativas longitudinales. Incluso autores tradicionales que han estudiado en profundidad el análisis del cambio (por ejemplo, Heck & Thomas, 2000; Little, Schnabel & Baumert, 2000), ni siquiera mencionan dicho efecto.

En definitiva, sigue existiendo una serie de cuestiones importantes en torno a dicho fenómeno, que no se han resuelto de forma clara hasta el momento, por ejemplo, ¿qué es el ERM?, ¿a qué se debe realmente el efecto de regresión a la media?, ¿se mantiene el ERM en los diseños con más de dos ocasiones de medida?, ¿el ERM es un artefacto estadístico o realmente se encuentra en los datos y, por tanto, no es necesario tratarlo, ya que forma parte de la naturaleza de los mismos? Todas estas cuestiones serán contestadas a lo largo de dicho capítulo, ya que, a pesar de la necesidad de estudiar el ERM, hasta el momento no se han encontrado respuestas a todas estas preguntas.

IV.1.1. ¿Qué es el efecto de regresión a la media?

La regresión hacia la media es un concepto que se invoca a menudo de forma simplista en la discusión del cambio, lo que conlleva que siga habiendo una gran necesidad de explicarlo con más detalle. Al hacerlo, se identifican las condiciones bajo las cuales el efecto de regresión a la media puede estar presente y, también aquellas donde estos efectos son minimizados, inexistentes o manifestados de manera que están en desacuerdo con las expectativas tradicionales.

El concepto de Regresión hacia la Media fue descubierto por Galton (1886) cuando estudiaba la relación entre la altura de los padres y la de sus hijos. Galton, se dio cuenta que los padres con una altura superior a la media, tenían hijos que también tendían a ser más altos que el promedio, pero se acercaban más a la altura media que sus padres. Lo mismo observó con los padres que tenían una altura inferior a la media, sus hijos tenían una altura que se encontraba por debajo del promedio, pero más cerca de la altura media de la población. Dicho fenómeno se refiere al hecho de que, en presencia de errores de medición, unidades experimentales que dan lugar a valores extremos en una primera medición, producen valores menos extremos en una segunda toma de muestras

independientes (o repetición). Este cambio de los valores extremos hacia la media de la población se produce a pesar de que no ha habido ninguna intervención o cambio en las distribuciones de la población o de errores subyacentes.

Dicho fenómeno fue denominado por Galton *regression toward mediocrity,* cuya causa fundamental se debe a la falta de fiabilidad de los instrumentos y al error de medición (Galton, 1886).

La regresión hacia la media se considera en el establecimiento de al menos dos ocasiones de observación, que implican la misma variable de medida y las mismas unidades experimentales, o bien dos variables relacionadas en una sola ocasión (por ejemplo, peso y altura).

En la literatura, el artículo de Furby (1973) es quizás uno de los mayores esfuerzos por tratar de explicar el fenómeno de la regresión hacia la media. Furby explicó el efecto de regresión de la siguiente manera. Considérese un grupo de sujetos para quienes se tiene información sobre dos variables, x e y. Con propósitos ilustrativos, se operacionalizan estas variables como, la cantidad de agresión observada en niños de guardería en una situación experimental estándar. Sea x la cantidad de agresión mostrada por un sujeto en el Tiempo 1 (medida mediante técnicas de observación del comportamiento), y sea y la cantidad de agresión mostrada por el mismo sujeto en un tiempo posterior (Tiempo 2). $\overline{X}$ representa el nivel medio de agresión para todo el grupo en el Tiempo 1, e $\bar{Y}$ representa el nivel medio de agresión del grupo en el Tiempo 2. Se parte de que: (a) las medidas de agresión son perfectamente fiables (sin error) (por ejemplo, dos observadores concuerdan perfectamente al observar el mismo sujeto en la misma ocasión) y (b) la correlación entre x e y es menor que 1; para simplificar la presentación, supongase además que (c) las distribuciones de las dos variables son normales y (d) las desviaciones estándar de las dos variables son iguales (es decir, $S_x = S_y$). Esta situación se representa gráficamente en la Figura IV.1, donde cada sujeto puede ubicarse según sus puntajes x e y. La elipse encierra todos los casos en el diagrama de dispersión (es decir, todos los sujetos), la línea continua azul representa la línea mejor ajustada para y regresado en x (cuya pendiente es r_{xy}, la correlación entre x e y), y la línea discontinua, representa $x = y$ (es decir, la línea de regresión si hubiera una correlación perfecta entre x e y).

Si ahora se selecciona un cierto nivel de agresión en el Tiempo 1 (cualquier nivel que no sea la media) y se denomina x', y a continuación, se seleccionan aquellos sujetos cuya puntuación x es x', y se calcula su puntuación media en y la cual se llama y' (es decir, para todos los sujetos que muestran x', la cantidad de agresión en el Tiempo 1 se calcula como la cantidad media de agresión mostrada en el Tiempo 2). Se encontrará que el nivel medio de agresión en el Tiempo 2 para estos sujetos, esta realmente más cerca del nivel de agresión promedio grupal en el Tiempo 2, que el nivel de agresión obtenido en el Tiempo 1, en comparación con la media general del grupo en ese mismo momento temporal. Esto se representa en la Figura IV.1, por el hecho de que x 'está más lejos de $\overline{X}$ que y' es de $\bar{Y}$. Si no hubiera efecto de regresión a la media las puntuaciones en y habrían sido y*, la cual esta tan lejos de $\bar{Y}$ como x' lo está de $\overline{X}$.

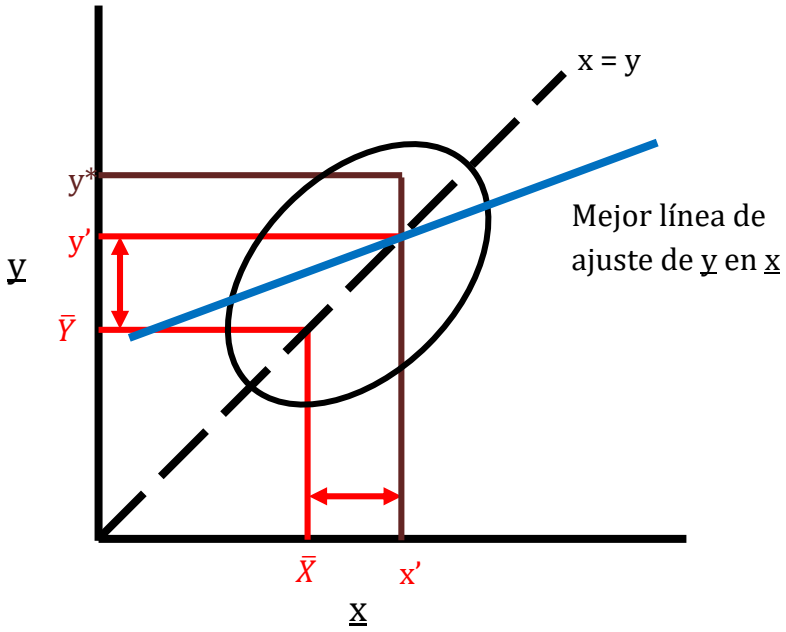


Figura IV.1. *Representación del efecto de regresión a la media*
Fuente: elaboración propia a partir de (Furby, 1973)

Este fenómeno se conoce como "regresión hacia la media" porque x e y a menudo representan la misma variable operativa (en el ejemplo, ambas son medidas de agresión observada), pero se miden en diferentes ocasiones. Los sujetos con un determinado nivel de agresión x' parecen "regresar hacia la media" cuando se mide un poco más adelante en esa misma variable. Otra forma de ver esto gráficamente se representa en la Figura IV.2, que contiene la misma gráfica de dispersión que en la Figura IV.1. Todos los sujetos cuyo nivel de agresión inicial era

x', pueden encontrarse en algún lugar a lo largo de la línea azul continua, ya que todos los sujetos están contenidos en la elipse. Las puntuaciones de estos sujetos y se pueden encontrar dentro de la proyección de la línea continua gruesa en el eje $\underline{y}$, y su puntuación media y es y'; y 'esta más cerca de $\bar{Y}$, de lo que x' esta de $\bar{X}$

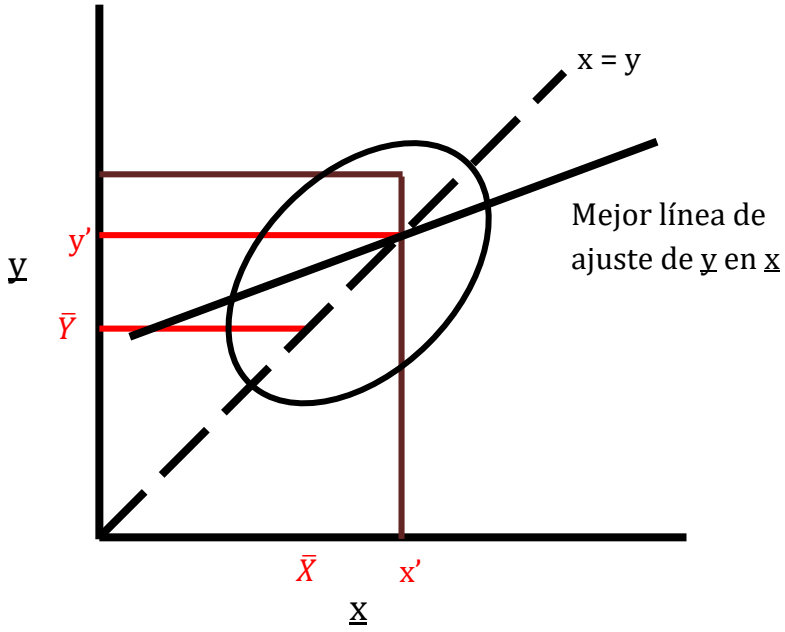


Figura IV.2. *Rango de puntaciones y para todos los sujetos cuya puntuación x es x'*
Fuente: elaboración propia a partir de (Furby, 1973)

Aunque en el ejemplo anterior para simplificar la presentación, se ha supuesto que las desviaciones estándar son iguales, el efecto de regresión aún se obtiene cuando $S_x \neq S_y$ y el caso más general puede expresarse de la siguiente manera: para una puntuación dado en x (por ejemplo, x'), la media correspondiente a la puntuación en y (por ejemplo, y') está más cerca de $\bar{Y}$ en unidades de desviación estándar de lo que x' esta de $\bar{X}$ en esas mismas unidades (Furby, 1973).

En resumen, una correlación imperfecta entre x e y conduce a efectos de regresión. Dicha correlación, puede deberse a la falta de fiabilidad en x e y, a un cambio diferencial real en los sujetos, o a ambos.

En la mayoría de los estudios de educación, los instrumentos de medición no son totalmente fiables y, además, se suele contar con dos variables separadas por un período de tiempo significativo, que no están perfectamente correlacionadas entre sí. Por lo tanto, es esperable siguiendo a Furby (1973), una regresión hacia la media tanto del error de medición, como de los cambios reales de los sujetos. Pero, tal y

como se estudiará en los siguientes apartados, esta situación solo es aplicable en el caso de dos ocasiones, la generalización a múltiples mediciones, cuando se siguen a los mismos sujetos implica connotaciones adicionales.

IV.1.2. ¿Por qué se produce el efecto de regresión a la media?

Se han utilizado dos líneas principales de razonamiento para explicar por qué se produce una falta de correlación perfecta y regresión hacia la media. La primera se refiere a la naturaleza de los errores de medición y la segunda se encuentra vinculada a la identificación de factores causales que operan entre ocasiones (Nesselroade et al.,1980).

IV.1.2.1. El error de medida

Este argumento, expresado en la Teoría Clásica de los Test (TCT) y en la situación de dos ocasiones de medida, depende de la suposición de que los errores de medición no están correlacionados con las mediciones, y que en la primera ocasión los individuos situados en grupos extremos (puntuaciones altas o bajas), tienen una mayor probabilidad de cometer errores positivos (altos) o negativos (bajos) en sus puntuaciones observadas, que en la segunda ocasión.

Los sujetos con un alto error positivo en la primera medición de una variable determinada, es probable que tengan mayores puntuaciones en la siguiente aplicación, en términos promedios, que aquellos sujetos que tienen altos errores negativos. Sin embargo, es poco probable que estos sujetos con un alto error positivo en la primera ocasión, también, tengan un alto error positivo en la segunda medición (ya que, los errores no están correlacionados a lo largo de las medidas). Por lo tanto, las puntuaciones de estos sujetos en la segunda ocasión tienden a ser más bajas (más cercanas a la media) que sus puntuaciones en la primera aplicación (la afirmación análoga pero opuesta es verdadera para un error negativo alto en la primera ocasión). En este sentido, en el caso de dos ocasiones de medida y en términos de puntuaciones observadas, en la segunda aplicación los grupos extremos se están "moviendo" más cerca de la media general que en la primera medida. Dicha interpretación supone que este movimiento no se encuentra en las puntuaciones verdaderas de los sujetos (que se mantienen constantes a lo largo del tiempo), sino que se produce en las puntuaciones observadas debido al error de medida. Siendo

la magnitud de dicho fenómeno mayor cuanto menos fiable sea el conjunto de mediciones realizadas (Cook & Campbell 1979; Ladd & Walsh, 2002; Rogosa, 1995).

Para entender mejor esta situación, imagínese una población de individuos en la cual se mide alguna característica métrica y se ajusta a una distribución normal con una cierta media y varianza. Siempre que no haya ningún cambio en los errores sistemáticos (en comparación con la medición), si la población es reevaluada, la media y la variación permanecerán cerca de la medida inicial. Sin embargo, esto obscurece el hecho de que el componente de error de las mediciones ha cambiado. Para los individuos cerca de la media de la población, algunos términos de error se han vuelto más grandes y otros más pequeños, lo que resulta una dispersión proporcional de los valores alrededor de la media original.

Un individuo que puntuó en el extremo más alto en la primera medida por tener un gran valor real, o por tener un error que fue muy positivo, obteniendo una puntuación muy alta (de manera similar para las puntuaciones más bajas). Si estos individuos fueran reevaluados, es improbable que tengan un error positivo tan grande en la segunda medición, de ahí que en esta segunda medida el sujeto obtenga una puntuación inferior a la primera, encontrándose más cerca de la media global si su puntuación se debe al error. Esto ocurre porque el error de medida es independiente entre las aplicaciones y, porque según aumenta el número de mediciones, las puntuaciones se ven arrastradas hacia la puntuación media, debido a que en el centro de la distribución se encuentra un mayor número de casos, de esta manera los errores extremos se estabilizan y las puntuaciones tienden hacia la media real (Figura IV.3).

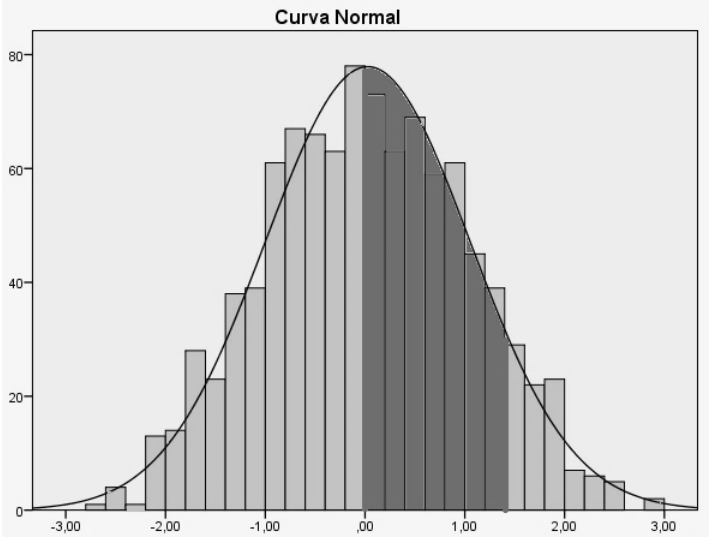


Figura IV.3. *Representación de casos en una distribución normal*
Fuente: elaboración propia

Inicialmente, si uno está dispuesto a aceptar los principios básicos de la teoría clásica de la medición que están involucrados, el razonamiento anterior es correcto. Sin embargo, como se demuestra más adelante, el argumento es aplicable sólo a la situación en la cual se realizan dos mediciones, seleccionando a los sujetos una única vez.

Como se ha señalado en este mismo capítulo, el error de medición es un indicador fundamental de la existencia del efecto de regresión a la media. Por lo tanto, es importante explorar en profundidad este concepto.

Cualquier medida cuantitativa es la suma de dos valores: el valor verdadero y el error de medición, tal y como se puede observar en la ecuación siguiente:

$$X_i = S_i + e_{ij} \qquad \textbf{Ec.IV.1}$$

Donde la puntuación observada (X) del sujeto i, esta formada por la puntuación verdadera (S) de dicho individuo i y e ,que es el error en la medición de esa puntuación verdadera, en ese mismo sujeto (i), en cada ocasión de medida (j). En las pruebas educativas, la "puntuación verdadera" es el valor estadístico esperado de la calificación de la prueba de una persona (Nesselroade et al., 1980). La diferencia entre una calificación de la prueba observada y la puntuación verdadera es el error de medida.

La medición precisa de una determinada variable es siempre deseable, pero todos los resultados de las pruebas contienen error de medición, es decir, la puntuación de la prueba obtenida no es probable que sea igual a la puntuación verdadera de la persona. La puntuación verdadera de una persona en el momento de la prueba, sólo puede tener un valor y por lo tanto es fijo. Sin embargo, la puntuación de la prueba obtenida de la persona, es una muestra de una distribución de una variable aleatoria, debido al error. Esta distribución específica permite otorgar una probabilidad asociada a cada valor posible obtenido en una prueba.

Por definición, la medida específica de una variable aleatoria variará, pero la probabilidad de obtener un cierto valor se puede determinar cuando la distribución de la variable es conocida. La puntuación verdadera y la varianza del error de medición son los dos parámetros necesarios para la distribución de las puntuaciones obtenidas, suponiendo normalidad.

La fiabilidad típicamente se ha calculado para conocer qué proporción de la varianza de la puntuación obtenida se debe a la varianza verdadera. Para ello, tradicionalmente se ha venido utilizando el coeficiente de correlación entre medidas repetidas o formas paralelas (ver ecuación IV.2), ya que permite estudiar si se replican las mismas medidas, al aplicar una prueba paralela, con ítems distintos. Si ambas formas son paralelas, la correlación entre ambas indica su grado de equivalencia, a mayor correlación más precisión entre las formas. (Abad, Olea, Ponsoda & García, 2011).

$$\rho_{xv}^2 = \frac{\sigma_v^2}{\sigma_x^2} = \rho_{xx} \qquad \textbf{Ec.IV.2}$$

El coeficiente de fiabilidad p_{xx}, es matemáticamente igual al cociente entre la varianza de las puntuaciones verdaderas (σ_V^2), y la varianza de las puntuaciones empíricas (σ_X^2) de cualquiera de las formas paralelas.

IV.1.2.2. Eventos causales

Una segunda explicación para la falta de correlación perfecta y la regresión, tiene que ver con la identificación de factores causales que operan entre ocasiones. Por ejemplo, Clarke, Neill y Welbourn (1960) y Furby (1973) reconocieron que las puntuaciones extremas (además de la independencia de los componentes del error de medición) en una distribución pueden reflejar, al menos en parte, combinaciones

relativamente raras de eventos. Estas combinaciones extrañas de acontecimientos tienden a no ser mantenidas en el tiempo y, por lo tanto, los individuos expuestos en una ocasión no lo estarán posteriormente, mostrando de esta manera un cambio hacia la media general. Dicho argumento es en muchos aspectos una variación del error de medida visto, aunque trata de los antecedentes del cambio en lugar de su medición. Además, el mismo es una posible explicación pero no necesaria, ya que son posibles concepciones alternativas de factores causales. Por ejemplo, las aplicaciones de factores causales no implican necesariamente eventos aleatorios y pueden conducir a la regresión.

IV.1.3. Especificaciones del efecto de regresión a la media en los diseños longitudinales

Hasta ahora, dentro del apartado IV.1 se ha estudiado de manera general el concepto de regresión a la media y las causas del mismo, pero es necesario profundizar más en dicho fenómeno con el fin de conocer su comportamiento en los diseños longitudinales. Para ello, previamente conviene revisar una serie de supuestos que muchas veces se asumen de manera implícita a la hora de explicar el ERM.

Tanto la anotación, así como la fundamentación utilizada para la explicación del estudio del efecto de regresión a la media en los diseños longitudinales, se realiza tomando como referencia el artículo de Nesselroade et al. (1980). Dichos autores, intentan explicar en profundidad el ERM en un modelo longitudinal de dos niveles. Sin embargo, en los siguientes apartados (IV.1.3.1., IV.1.3.2. y IV.1.3.3), se amplian las explicaciones realizadas por dichos autores acerca del ERM, al caso de los modelos jerárquicos lineales con tres niveles (tiempo, estudiante y aula), ya que son desarrollados en la parte metodológica de este trabajo.

En la explicación sobre el ERM que se realiza de aquí en adelante, se consideran que las pruebas han sido escaladas, y que las desviaciones estándar de las puntuaciones en una ocasión son las mismas que en otra medición. Para simplificar la notación, esta desviación estándar común se toma como la unidad, y los resultados de las pruebas como desviaciones de la media de la población. Por lo tanto, si un sujeto es seleccionado aleatoriamente de una población, su puntuación en la ocasión i tendrá un valor esperado de $E(X_1) = 0$ y la desviación tipica es

desviación estándar σ(X_i) = 1, para cada medida i. Así mismo se considera que los sujetos seleccionados en la ocasión 1 tienen una puntuación x ($X_1 = x$).

Además, se mantienen una serie de supuestos de la Teoría Clásica de los Test (TCT) y las consecuencias derivadas de los mismos (Abad et al., 2011):

a) El modelo clásico se basa en el concepto de puntuación observada (X_i) compuesta, tal y como se ha estudiado en este capítulo, por la puntuación verdadera (S_i) y el error de medida (e_i). Para una persona i, la puntuación verdadera es constante pero las puntuaciones observadas y los errores son variables aleatorias.
b) Las puntuaciones verdaderas y el error para una población examinada no están correlacionadas: ρ(S, e_i) = 0. Los errores no están correlacionados con algún atributo de la persona , ya que de lo contrario no serían aleatorios.
c) El valor verdadero y el error de medición en tests distintos no correlacionan: $\rho(\mathrm{S}, e_{iTEST1,}\ \mathrm{S}, e_{iTEST2}) = 0$.
d) Los valores de error en tests distintos no correlacionan: $\rho(e_{i_{TEST1}}, e_{i_{TEST2}}) = 0$.
e) El valor esperado del error de medida es igual a 0: $E(e_i) = 0$.
f) La relación entre una variable aleatoria (X) y su error (e_i), es aditiva dando lugar a un modelo lineal: $E(x|\mathrm{e}\ ) = a + b * x_1$ .Donde a vale cero y b es la correlación entre x_1 y x_2.
g) La covarianza de todas las ocasiones de medida es 0: $cov\ (X_i = 0)$.
h) La varianza de las puntuaciones es igual a 1 y es igual a la varianza de la puntuación verdadera, más la varianza del error: $\sigma^2(x_i) = 1 = \sigma^2(S) + \sigma^2(e_i)$.
i) La covarianza de dos puntuaciones observables es igual a la varianza de la puntuación verdadera: $cov\ (X_i, X_j) = \sigma^2(S)$.

IV.1.3.1. El efecto de regresión a la media en los diseños con dos ocasiones de medida

La mayor parte de la discusión sobre el ERM se ha centrado en el estudio de dos variables o de una variable medida en dos momentos distintos (modelos de ganancia), ya que ha llegado a ser conocida como una de las clásicas amenazas a la validez interna en los diseños pre-test-post-test (Cook & Campbell, 1979),

produciéndose la regresión de un pre-test sobre un post-test cuando las variables se correlacionan de manera imperfecta (Doran, 2003).

Estas situaciones donde el número de mediciones es limitado, maximizan el análisis de los problemas asociados con la regresión hacia la media (Rogosa, 1995). Sin embargo, el ERM se puede modificar cuando se tienen en cuenta más de dos ocasiones de medida, y los sujetos evaluados son los mismos a lo largo de las diferentes pruebas.

Nesselroade et al. (1980) consideran que el fenómeno de regresión hacia la media en dos ocasiones (puntuaciones X_1 y X_2), se refiere a la esperanza condicional de X_2, dado que la prueba se administró en la ocasión 1 y se ha registrado la puntuación como $X_1 = x$.

Bajo la suposición, tal y como se ha estudiado anteriormente (ver ecuación IV.1) de que la puntuación observada (X_i) esta formada, por la puntuación verdadera del sujeto (S_i), más el error específico de la ocasión de medida (e_{ij}), siendo estos errores aleatorios e independientes con media cero y varianza común ($\mathrm{e_{tij}} \sim \mathrm{N}(0, \sigma^2)$). Resulta que tal y como estan definidos los valores tipificados de X_1 (pre-test) y X_2 (post-test) se consideran medidas paralelas, ya que la puntuación verdadera que miden es la misma en las distintas ocasiones, y aunque el error es un valor distinto, procede de la misma distribución, es decir, con media cero y varianza común, por lo tanto:

Ec. IV.3

$$X_1 = S + e_1 \qquad \mathrm{e}_1 \sim \mathrm{N}(0, \sigma_e^2)$$

$$X_2 = S + e_2 \qquad \mathrm{e}_2 \sim \mathrm{N}(0, \sigma_e^2)$$

Luego las varianzas son iguales en las medidas

$$\sigma_{e1}^2 = \sigma_{e2}^2$$ **Ec. IV.4**

Así, ocurre lo mismo con las varianzas de cada ocasión que están compuestas por la varianza de la puntuación verdadera $\sigma^2(S)$ y la varianza del error $\sigma^2(e)$:

Ec. IV.5

$$\sigma^2(X_1) = \sigma^2(S) + \sigma^2(e)$$

$$\sigma^2(X_2) = \sigma^2(S) + \sigma^2(e)$$

Por lo tanto los componentes de la varianza en cada ocasión son las misma:

$$\sigma^2(X_1) = \sigma^2(X_2)$$ **Ec. IV.6**

Y también las desviaciones típicas:

$$\sigma(X_1) = \sigma(X_2)$$ **Ec. IV.7**

Al multiplicar las desviaciones se obtiene la varianza de la puntuación observada:

$$\sigma(X_1) * \sigma(X_2) = \sigma^2(X)$$ **Ec. IV.8**

Por lo tanto ambas ocasiones de medidas paralelas suponen:

Ec. IV.9

$$\rho\ (X_1, X_2) = \frac{cov(X_1, X_2)}{\sigma(X_1) * \sigma(X_2)} = \frac{\sigma^2(S)}{\sigma^2(X)} = \rho_{SX}^2 \text{ (coeficiente de fiabilidad de la medida)}$$

Cuando X_1 y X_2 están tipificados como ocurre en este caso ($\sigma(X) = 1$,) entonces la correlación (ρ) entre la primera ocasión de medida y la segunda es igual a la varianza de la puntuación verdadera ($\rho\ (X_1, X_2) = \sigma^2(S)$).

Así, si se estudia a un grupo de sujetos cuya puntuación en la ocasión 1 es x, es decir, $X_1 = x$, la puntuación esperada de estos sujetos para la ocasión 2 (X_2) es $E(X_2|X_1 = x)$, y como la correlación entre la ocasión 1 (X_1) y 2 (X_2) es $\rho(X_1, X_2) = \sigma^2(S)$, entonces la esperanza condicional de X_2 viene dada por:

$$E(X_2|X_1 = x) = \rho\ (X_1, X_2) * x = \sigma^2(S) * x$$ **Ec. IV.10**

Ya que la $\sigma^2(S)$ es igual a la correlación (X_1, X_2) y esta suele ser siempre menor a 1, si X_1 y X_2 son medidos en unidades de sus desviaciones estándar, entonces dado $X_1 = x$, se espera que X_2 esté más cerca de su media $[E(X_2) = 0]$, haciendo que de la primera ocasión de medida a la segunda se produzca una regresión hacia la media.

Por ende, si se selecciona a un número muy grande de individuos con puntuaciones cercanas a x en la ocasión 1, su puntuación media en la ocasión 2 será muy cerca de la $\rho\ (X_1, X_2) * x$. Algunos sujetos habrán puntuado más bajo, algunos más altos, y algunos incluso más altos que en la puntuación inicial (x), pero el promedio será muy cercano a la esperanza ($\rho\ (X_1, X_2) * x$). Esta situación es la que se ha mostrado en la figura IV.1.

A la vista de lo comentado hasta el momento en este apartado, la falta de correlación perfecta entre las puntuaciones, hace que el ERM sea algo inevitable en los modelos de ganancia con dos puntuaciones, lo que debe tenerse en cuenta, sobre todo, cuando se tomen decisiones de alto impacto basadas en los resultados de este tipo de modelos.

IV.1.3.2 El efecto de regresión a la media en los diseños con más de dos ocasiones de medida.

Es importante entender que la regresión hacia la media, no es un fenómeno omnipresente, ni es siempre continuo a través de las ocasiones de medida.

Cuando se realizan más de dos mediciones en los mismos sujetos, es una *falacia* considerar la comparación por pares, tal y como se muestra en la siguiente ecuación IV.11:

Ec. IV.11

$$E(X_2|X_1 = x) = \rho\ (X_1, X_2) * x,$$

$$E(X_3|X_2 = y) = \rho\ (X_2, X_3) * y,$$

$$E(X_4|X_3 = z) = \rho\ (X_3, X_4) * z,$$

Aunque estas ecuaciones son correctas para dos ocasiones de medida, conducen a una conclusión equivocada en la presente situación de múltiples mediciones, ya que dicha formulación supone seleccionar a los individuos en distintos momentos cada vez, y no realizar un seguimiento de los mismos sujetos a lo largo del tiempo. Es decir, el malentendido reside en imaginar una secuencia esperada de resultados de las pruebas, empezando por $X_1 = x$, y acercándose sucesivamente a la media de cero, como si la regresión hacia la media implicara una regresión que continúa más allá de la segunda ocasión, como a menudo se asume.

Intuitivamente, dado $X_1 = x$, se podría ingenuamente esperar la siguiente secuencia:

Ec. IV.12

$$X_1 = x,$$

$$X_2 = \rho\ (X_1, X_2) * x,$$

$$X_3 = \rho\ (X_2, X_3) * X_2, = \rho\ (X_2, X_3) *\ corr\ (X_1, X_2) * x$$

$$X_4 = \rho\ (X_3, X_4) * X_3, = \rho\ (X_3, X_4) *\ corr\ (X_2, X_3) *\ \rho\ (X_1, X_2) * x$$

Así, por ejemplo, si $\rho\ (X_i, X_{i+5}) = 0{,}5$ para todo i y $X_1 = 2{,}0$, la secuencia esperada seria: $X_1 = 2{,}0,\ X_1 = 1, X_1 = 0{,}5,\ X_1 = 0{,}25,$ etc.

De acuerdo con esta serie, parece que la regresión a la media se arrastra de una ocasión de medida a otra. Por ello, cuando se trabaja con más de dos mediciones es un error considerar dicha formulación.

Este proceso estocástico se conoce como secuencia de Markov, ya que el estado siguiente de una variable depende del estado actual, así en el estado actual está resumida toda la información (Frederiksen & Rotondo, 1979; Glass, Willson & Gottman, 1972; Nelson, 1973). Por lo tanto, si se desea calcular $E(X_3|X_1\ = x,\ X_2 = y\ )$ acorde a la secuencia de Markov bastaría con calcular, $E(X_3|X_2\ = y\ )$, ya que: $E(X_3|X_2\ = y\ ) = E(X_3|X_1\ = x,\ X_2 = y\ )$.

Sin embargo, cuando se trabaja con medidas repetidas, hay que tener en cuenta que un individuo se selecciona solo una única vez, teniendo una puntuación determinada ($X_1 = x$)en la primera medición, de tal manera, que para generar la secuencia futura es necesario tener en cuenta dos factores: el valor de la puntuación seleccionada o el criterio de selección $X_1 = x$, y las características de la población de las puntuaciones observadas (incluyendo la media de la población, $E(X_1) = 0$ y la $corr\ (X_{1,}X_i)$).

Para entender mejor cómo funciona el ERM con más de dos medidas, supongan que se selecciona en la ocasión 1 a un grupo de sujetos cuyo valor en $X_1 = x$. Existiendo dos sujetos en ese grupo (el primero y el segundo, por ejemplo) que tienen la misma puntuación observada, siendo esta $x = 10$, pero su puntuación verdadera es distinta de 10. Por ejemplo, el sujeto número 1 seleccionado tiene una puntuación verdadera (S_i) de 7 y su error (e_i) es de 3, el sujeto número 2 también tiene una puntuación observada de 10 pero su puntuación verdadera es de 4 y su error de 6. En esta situación, la puntuación observada es la misma para los dos sujetos, aunque cuenten con distinta puntuación verdadera $(\bar{S}|x = 10) \neq 10$. Esto es debido a que el error en ambos casos es distinto de 0 $(\bar{e}|x = 10) \neq 0$, teniendo a veces un efecto positivo y otros negativo, y a que la suma de la puntuación observada más el error dan el mismo resultado $((\bar{S}|x = 10) +\ (\bar{e}|x = 10) = 10\ )$. Al no ser el error 0, la puntuación verdadera estimada, no puede ser 10, y siempre que el

número de sujetos no sea infinito esto ocurrirá, ya que el valor del error será siempre distinto de 0, y como consecuencia la puntuación observada no coincidirá con la puntuación real. De esta manera se puede encontrar, por ejemplo, a otro sujeto donde su puntuación observada también es de 10 ($\bar{S}|X_1 = 10$), pero su error es de 1 ($\bar{e}|X_1 = 10$), en este caso la puntuación verdadera que le corresponde a dicho sujeto sería de 9.

Siguiendo el razonamiento utilizado para la explicación del ERM con dos ocasiones de medida (*apartado IV.1.3.1.*), el valor de la puntuación observada en la segunda ocasión vendría definido como:

$$E(X_2|X_1 = x) = \rho\,(X_1, X_2) * x \qquad \textbf{Ec. IV.13}$$

En la tercera ocasión, si se aplicara la secuencia de Markov, se establecería una falacia debido a que se procedería de la siguiente manera:

$$E(X_3|X_2 = x) = \rho\,(X_1, X_2) * x = \rho\,(X_2, X_3) * = \rho\,(X_1, X_3) * x \qquad \textbf{Ec. IV.14}$$

Y así sucesivamente con el resto de ocasiones.

Pero dicho razonamiento es incorrecto, ya que la secuencia anterior supone que se selecciona en cada ocasión a sujetos distintos. Cuando en realidad se seleccionan una única vez (generalmente en la primera ocasión) y se van siguiendo a los mismos a lo largo del tiempo. Estos sujetos cuyo valor en $X_1 = x$, tienen como media su puntuación verdadera (S), ($\bar{S}|X_1 = x$), entonces, y dado que su puntuación verdadera no ha cambiado de una aplicación a otra, ya que son los mismos sujetos seleccionados de manera aleatoria en la primera aplicación, la ecuación correcta para la tercera aplicación sería la siguiente:

$$E(X_3|X_1 = x) = (\bar{S}|X_1 = x) \qquad \textbf{Ec. IV.15}$$

Y lo mismo para el resto de ocasiones:

$$E\left(X_j\middle|X_1 = x\right) = (\bar{S}|X_1 = x) \qquad \textbf{Ec. IV.16}$$

Es decir, el valor esperado en la ocasión j de los sujetos que en la ocasión 1 tuvieron $X_1 = x$, es siempre la media de la puntuación verdadera (S) de esos mismos sujetos, es decir, la media en la puntuación verdadera de los sujetos que en la primera ocasión, siguiendo con el ejemplo anterior, tuvieron una puntuación de 10.

De esta manera, en el ejemplo en el que un sujeto tenía una puntuación observada de 10 y un error de 1, el valor esperado en la segunda medición de este sujeto que obtuvo $x = 10$ en la primera ocasión, sería de 9, ya que el error vale 1. Como es el mismo sujeto a lo largo de todas las mediciones, el valor esperado para cada una de las puntuaciones es siempre 9, por lo que no habría regresión a la media más allá de la segunda ocasión de medida.

Sólo hay regresión a la media entre la primera y la segunda ocasión, debido a que los sujetos han sido seleccionados por su valor en X_1 que contiene error, y no por su valor en S (puntuación verdadera) que es desconocido. Por lo tanto, si se conociera el valor de S de los sujetos y se seleccionaran por este valor, no habría regresión a la media ni si quiera en el paso de la primera a la segunda medición.

Así el punto clave está en seguir longitudinalmente el mismo grupo de individuos seleccionado de manera aleatoria en la primera ocasión de medida. Esto significa que el criterio de selección $X_1 = x$ permanece en todas las medidas, ya que sólo se selecciona a los sujetos en un único momento con una única puntuación verdadera que se mantiene constante. Por lo tanto, la secuencia esperada debe tener siempre presente la ocasión en la que se produjo la selección (generalmente, la primera medición). Así, la secuencia esperada y correcta para cada una de las diferentes ocasiones sería la siguiente:

Ec. IV.17

$$E\ (X_2|X_1 = x),$$

$$E\ (X_3|X_1 = x),$$

$$E\ (X_4|X_1 = x),$$

Mientras que la secuencia incorrecta, sería:

Ec. IV.18

$$x$$

$$\rho\ (X_1X_2) * x,$$

$$\rho\ (X_2X_3) * \rho\ (X_1X_2) * x$$

$$\rho(X_3X_4) * \rho\ (X_2X_3) * \ \rho\ (X_1X_2) * x$$

La principal diferencia existente entre ambas secuencias, es que en el caso de la primera siempre se considera el mismo grupo de individuos (los que se seleccionan en la primera ocasión de medida) mientras que, en la segunda secuencia (secuencia de Markov) se seleccionan a los individuos que tienen una determinada puntuación en el paso 2 o 3 (dependiendo de la ocasión de medida referida), lo que significa que en cada paso se incorporan nuevos sujetos, y, por lo tanto, no se está realizando un seguimiento de los mismos, ya que no se tiene en cuenta la puntuación que los sujetos obtuvieron en la primera ocasión $E(X_2|X_1 = x) = corr\ (X_1, X_2) * x$, sino la que se obtuvo en las ocasiones anteriores $E(X_3|X_2 = y) = corr\ (X_2, X_3) * y$.

La secuencia esperada donde el ERM va desapareciendo a partir de la segunda ocasión de medida, se calcula asumiendo, que un grupo de individuos han sido seleccionados en la primera ocasión y tienen en dicha medida una puntuación $X_1 = x$. Si la selección se realiza en otra ocasión posterior, puede ocurrir una regresión adicional pero no siempre hacia la media general de cero. Por ejemplo, si todos los individuos seleccionados tienen puntuaciones $X_1 = x$ en la primera ocasión, y uno se selecciona después de la segunda ocasión teniendo puntuación $X_2 = 0$, se esperaría que su puntuación en la tercera ocasión regresara hacia la media inicial de la población seleccionada.

Aún así, Willet (1994,1997) considera que no siempre los cambios producidos en los agregados de las puntuaciones seleccionadas se pueden deber al ERM, sólo aquellos que puedan atribuirse a fluctuaciones aleatorias (errores de medida), el resto deben ser considerados como cambios en la puntuación real.

En conclusión, como en los diseños longitudinales, que se sigue al mismo grupo de sujetos, la regresión hacia la media no se extiende más allá de la segunda ocasión, es evidente que un control en el diseño es fácil. Si se seleccionan a los individuos en la primera ocasión, se logra una representación adecuada del cambio sin confundirse con los efectos de regresión, iniciando el trazado de las funciones de cambio en la segunda ocasión. En otras palabras, la primera ocasión sirve de criterio para seleccionar grupos extremos, y la segunda sirve como punto de iniciación para representar el cambio, que se acerca cada vez más a la puntuación verdadera del sujeto cuanto más cercana a 1 sea la correlación entre la puntuación del sujeto en dos ocasiones de medida diferentes. Lo que se desconoce hasta el momento, son el

número de mediciones necesarias para obtener estimaciones precisas (fiables), cercanas a la puntuación verdadera del estudiante, aspecto que será estudiado en la parte metodológica de dicha tesis.

IV.1.3.3. Tipos de patrones de correlación en el efecto de regresión a la media

La complejidad del ERM ha originado un gran número de falacias en torno a dicho concepto, y en torno a los posibles patrones de correlaciones existentes entre las puntuaciones de los sujetos a lo largo del tiempo.

En este sentido, se ha prestado mucha atención a la naturaleza de la estructura de los patrones de correlación subyacente en las puntuaciones longitudinales. Pero la discusión realizada hasta el momento del efecto de regresión a la media, no ha considerado suficientemente el papel de los diferentes patrones de correlación, y se ha asumido con frecuencia que las matrices de correlación simplex (patrón autorregresivo), son el patrón esperable en los diseños con medidas repetidas (Garson, 2013; Pardo & Ruiz, 2013; Raudenbush & Bryk, 2002; Ronald et al., 2010). Sin embargo, en los diseños longitudinales donde se sigue a los mismos sujetos, la forma de regresión hacia la media entre la primera medición y el resto, cambia significativamente según se asumen patrones alternativos de correlación entre las distintas ocasiones (Nesselroade, et al. 1980). De esta manera, el valor medio esperado vendrá determinado por el patrón de correlación entre la primera ocasión de medida y el resto de aplicaciones ($corr\ (X_1, X_i)$).

Obviamente, se puede concebir una gran variedad de patrones correlacionales, pero generalmente tres son los predominantes: (a) correlaciones constantes, (b) correlaciones que disminuyen con el aumento de la separación de observaciones en una secuencia (correlaciones simplex o autorregresivas), y (c) correlaciones que se incrementan al aumentar la separación de observaciones.

A) Correlaciones constantes

No necesariamente se debe incrementar el efecto de regresión a la media si la secuencia de correlación $\rho\ (X_1, X_i), i > 1$ es constante [por ejemplo, $\rho\ (X_1, X_2) = \rho\ (X_1, X_3) = \cdots = \rho\ (X_1, X_n) = 0{,}5$], en este caso se espera un tamaño del efecto de regresión invariable entre la ocasión 1 y 2, la ocasión 1 y 3, 1 y 4, y así sucesivamente.

Un ejemplo específico de este patrón de regresión particular está en el modelo del error de medición clásico. Supóngase que en más de *n* ocasiones de medida, las puntuaciones de un individuo seleccionado al azar, X_1, X_2, X_3, X_n estan compuestas, cada una de ellas, de una puntuación verdadera S que permanece sin cambios y un error de medida (e_i) tal que S, $e_1, e_{2,} \dots\ e_n$ que son mutuamente independientes (y, por lo tanto no correlacionados). Entonces $X_i = S + e_i$ donde i = 1,..., n y que $\sigma^2\,(X_i) = 1$ y $E(S) = 0 = E\,(e_i)$. Ahora $\sigma^2\,(X_i) = \sigma^2\,(S) + \sigma^2\,(e_i)$ [porque $cov\,(S, e_i) = 0$] y por lo tanto, $cov\left(e_i, e_j\right) = 0$ para i ≠ j, considerandose que $cov\left(X_i,\ X_j\right) = \sigma^2\,(S)$,y por lo tanto la $\rho\left(X_i, X_j\right) = \sigma^2\,(S)$ para i ≠ j. Se denota α = $\sigma^2\,(S)$, entonces 0≤ α ≤ (ya que α + $\sigma^2\,(e_i)$ = 1, es decir, la varianza total) y dado $X_1 = x$, la secuencia correcta esperada es: $x, \alpha x, \alpha x, \alpha x \dots, \alpha x$

Por lo tanto, como se muestra en la Figura IV.4, después del primer test, no se espera regresión adicional entre la ocasión 1 y el resto de ocasiones, sino que dicho fenómeno se mantiene constante. El efecto sobre los cambios en la puntuación esperada tras seleccionar al individuo teniendo puntuación $X_1 = x$, se disipa después de la segunda ocasión, pero el efecto de regresión entre la primera ocasión y el resto de ocasiones se mantiene invariante. Sólo cuando no hay error de medición α = 1 y por tanto la puntuación observada y la verdadera son las mismas, no hay efecto de regresión incluso desde la primera ocasión de medida.

La secuencia esperada incorrecta para este ejemplo habría sido $x, \alpha x, \alpha^2 x, \alpha^3 x$.. y así sucesivamente.

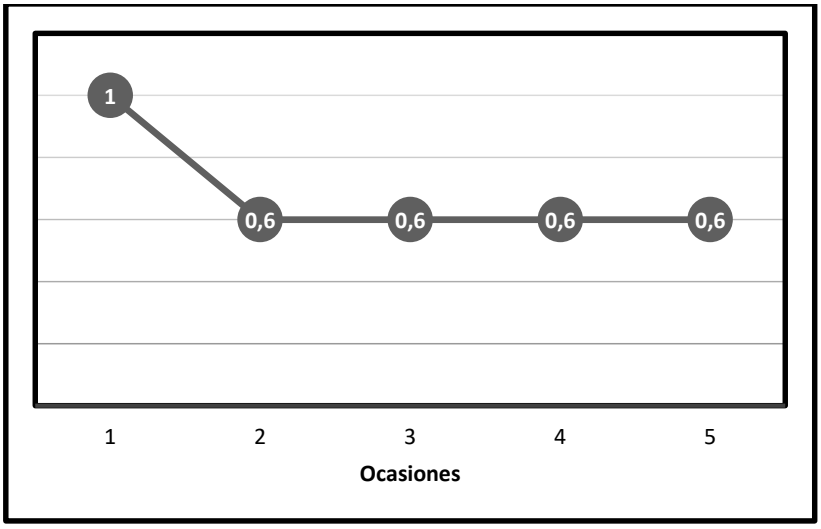


Figura IV.4. *Correlación constantes*
Fuente: elaboración propia a partir de Nesselrode et al., (1980)

Partiendo de las ideas anteriores, cuando se trabaja con datos multinivel (3 niveles: tiempo, estudiante y aula) y de carácter longitudinal, este patrón es esperable en situaciones en las cuales los estudiantes o aulas parten de los mismos niveles de logro, aunque su rendimiento a lo largo del tiempo sea diferente (cuando exista variabilidad en la pendiente y en el error intrasujeto, pero no en el estatus). También, en la situación contraria, en la cual las aulas o estudiantes parten de distintos niveles de rendimiento previo, aunque sus niveles de logro a lo largo del tiempo sean constantes, sin que exista mucha variabilidad en el error intrasujeto, es decir, en modelos donde exista varianza en el intercepto y en el error intrasujeto, pero no en la pendiente, en ambas situaciones es esperable un patrón constante. Si no existe variabilidad en el intercepto (u_{00j}), la puntuación para todas las aulas o estudiantes en la primera ocasión de medida es siempre la misma, por tanto, la correlación entre esta primera ocasión y el resto de ocasiones se esperaría que fuera constante. Lo mismo ocurriría si existiera varianza en el estatus y en el error intrasujeto, siendo la varianza del error intrasujeto muy pequeña en comparación con la de la varianza del intercepto, aquí, las correlaciones entre la primera ocasión de medida y el resto de mediciones también se esperarían que fueran constantes, porque siempre se estimaría una puntuación muy parecida, debido a que existiría muy poca variabilidad en el error intrasujeto. Esto se comprobará en la parte empírica del trabajo.

B) Correlaciones decrecientes

Un segundo patrón importante entre las puntuaciones de los sujetos en las distintas ocasiones de medida, es aquel en el que la secuencia de correlación $corr\ (X_1, X_i)$ decrece con el aumento de la distancia entre ocasiones. Este caso clásico de tal matriz de correlaciones es el denominado como matriz simplex o autorregresiva (Guttman, 1954).

En la Figura IV.5 se muestra la secuencia esperada correspondiente a este tipo de patrón, cuando $X_1 = 1$. Si $X_1 = x$ la secuencia de puntuaciones esperadas es multiplicado por x. La existencia de un patrón decreciente entre la ocasión 1 y el resto de ocasiones, supone que el efecto de regresión a la media se va incrementando según se alejan las mediciones de la ocasión 1 y, por lo tanto, que el error de medida es cada vez mayor, lo que supone un alejamiento de la puntuación verdadera.

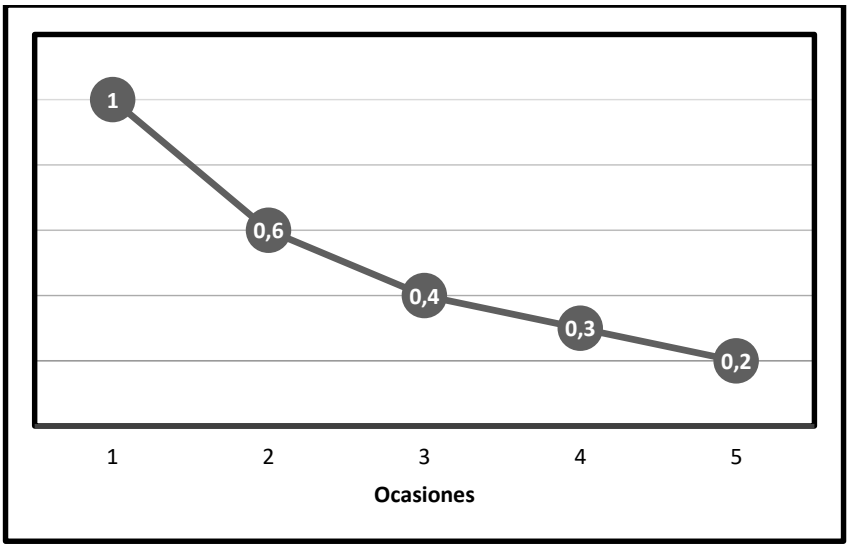


Figura IV. 5. *Correlación decreciente*
Fuente: elaboración propia a partir de Nesselrode et al. (1980)

Esta situación se da cuando por ejemplo el cambio de una ocasión a otra esta correlacionada con la media inicial. Por ejemplo, cuanto mayor es X_1 mayor es el cambio entre X_{k-1} y X_k, por lo tanto, en sujetos con valores grandes de $X_{1,}$el cambio acumulado al final será muy grande, mientras que, valores bajos de X_1 conlleva a cambios más pequeños.

El razonamiento matemático es el siguiente:

$$\sigma\left(X_{1,}\, X_k - X_{k-1}\right) = cov\left(X_{1,}\, X_k\right) - cov\left(X_{1,}, X_{k-1}\right) \qquad \textbf{Ec. IV.19}$$

Por lo tanto,

Ec. IV.20

$$E\left[(X_{1,} - E\left(X_{1,}\right))((X_K - X_{k-1}) - E\left(X_K - X_{k-1}\right))\right] =$$

$$E\left[(X_{1,} - E\left(X_{1,}\right))\left((X_K - X_{k-1}) - E\left(X_K\right) + E\left(X_{k-1}\right)\right)\right] =$$

$$E\left[(X_{1,} - E\left(X_{1,}\right))\left(X_K - E(X_k) - (X_{k-1}) + E(X_{k-1})\right)\right] =$$

$$E\left[\left(X_{1,} - E\left(X_{1,}\right)\right)[(X_K - E\left(X_k\right)) - [X_{k-1} - E(X_{k-1})]]\right] =$$

$$E\left[(X_{1,} - E\left(X_{1,}\right))(X_K - E\left(X_k\right))] - E[(X_1 - E(X_1)) - \left(X_{k-1} - E(X_{k-1})\right)]\right]$$

Es decir, la covarianza entre la primera medida $X_{1,}$ y el cambio entre dos medidas consecutivas posteriores, X_k y X_{k-1}, es la diferencia entre la covarianza de la primera medida (X_1) con la medida X_K.

Si se asume como en el caso anterior, que la correlación de X_1 con todas las demás medidas es constante, entonces la correlación de la primera medida con el cambio entre dos medidas consecutivas tiene que ser 0, ya que $cov(X_{1,} X_k - X_{k-1}) = cov\ (X_{1,} X_k) - cov\ (X_{1,}, X_{k-1})$. Si estas dos últimas covarianzas son iguales entonces $cov(X_{1,}\ X_k - X_{k-1}) = 0$.

Tradicionalmente, se ha considerado que este patrón es el más común cuando se trabaja con diseños longitudinales (Nesselrode, et al., 1980), quizás en parte debido al refuerzo sutil de la noción que proviene de la falacia de expectativas (falacia de Markov) discutida anteriormente.

En modelos de tres niveles (tiempo, estudiante y aula) y longitudinales, dicho patrón decreciente es esperable cuando la correlación entre el intercepto y la pendiente sea de intensidad baja positiva o de intensidad baja o alta negativa. Si la relación entre el rendimiento previo y el alcanzado por los estudiantes o las aulas a lo largo del tiempo es de intensidad baja, los niveles de logro de los estudiantes en la primera medición no estarán muy relacionados con el obtenido en el resto de mediciones y, por lo tanto, las correlaciones deberían tender a ser decrecientes. Del mismo modo, si la relación entre el intercepto y la pendiente es negativa, las puntuaciones de los sujetos, a medida que se alejan de la ocasión inicial serán cada vez más diferentes (el rendimiento de partida no estará relacionado con el que se alcanza a lo largo del tiempo), y, así, las correlaciones entre las puntuaciones de los estudiantes, según se alejan de la primera ocasión de medida, se espera que sean cada vez más pequeñas (decrezcan). Dicha comprobación se estudiará en la parte empírica de la tesis.

C) Correlaciones Crecientes

La tercera situación importante que se puede encontrar, es aquella en la que las correlaciones $corr\ (X_1, X_i)$, aumentan según las puntuaciones se alejan en el tiempo. En este caso el efecto de regresión va desapareciendo según se incrementa el número de mediciones, ya que las correlaciones tienden hacia 1.

Estas puntuaciones, cada vez se parecen más a la puntuación obtenida en la primera medición, puesto que se van apartando de la media, y regresan a la puntuación inicial, por ello, Nesselrode, et al. (1980) consideran que este patrón supone una egresión a la media.

En tal situación donde el patrón esperado es el creciente, si se utiliza la ocasión 1 como predictor para todas las ocasiones subsecuentes, se observa que la magnitud de la regresión disminuye continuamente en comparación con el efecto de la ocasión 1 a la ocasión 2 (ver figura IV.6). La secuencia de las puntuaciones esperadas para un individuo con la puntuación inicial $X_1 = x$ (la secuencia de correlación escalada esta multiplicada por x, por la regresión) disminuiría de la ocasión 1 a la ocasión 2, pero después de eso (a partir de la segunda medición), exhibiría egresión de la media, de tal manera que un sujeto con una puntuación inicial de $X_1 = x$, a partir de la segunda medida se aleja de la media (egresión a la media).

El cambio esperado entre las aplicaciones se modifica entre la ocasión k y k + 1. Un individuo seleccionado con una puntuación inicial $X_1 = x$, después de la regresión inicial que se da en sus puntuaciones entre la ocasión 1 y la ocasión 2, todos los cambios esperados futuros se alejan de la media, aumentando la correlación con la medida inicial.

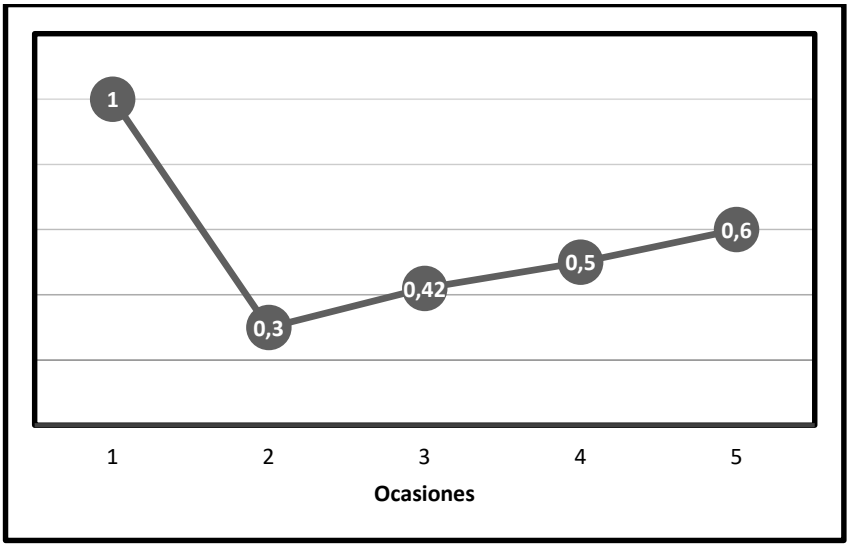


Figura IV.6. *Correlación creciente*
Fuente: elaboración propia a partir de Nesselrode et al. (1980)

Al contrario del patrón decreciente, en un diseño longitudinal de carácter multinivel, es esperable una correlación creciente a medida que las aplicaciones se alejan de la primera medición, cuando las puntuaciones de partida están muy relacionadas con las obtenidas a lo largo del tiempo, siendo dicho rendimiento alto en las diferentes ocasiones de medida, para las aulas y/o estudiantes que parten de niveles de rendimiento alto, y bajo para los estudiantes y/o aulas que parten de niveles de logro bajo, es decir, cuando exista una correlación de intensidad alta y de direccionalidad positiva entre el intercepto y la pendiente.

IV.1.4 ¿Cómo paliar el efecto de regresión a la media?

A pesar de que un mito ha perpetuado que uno no puede evitar la regresión a la media, considerándose como un artefacto omnipresente en la investigación (Furby, 1973). En la literatura existen ciertas sugerencias sobre cómo paliarlo, por ejemplo, Lord (1956) sugiere la utilización de una "puntuación de ganancia" basada en una técnica de ajuste de regresión. Campbell y Kenny (1999) ofrecen una serie de recomendaciones para evitar el efecto de regresión:

1. Recopilar datos de pre-test que proporcionen una estimación fiable de los efectos de selección (especialmente si los participantes son realmente asignados a diferentes grupos sobre la base de la puntuación pre-test).
2. Utilizar instrumentos que tengan fuertes propiedades psicométricas, reduciendo así el posible error de medición.
3. Comprender y recoger las medidas apropiadas para controlar las variables (incluida la autoselección), utilizadas para asignar a los participantes a grupos no equivalentes.
4. Explorar resultados gráficos de los datos obtenidos.

Roberts (1980), ha desarrollado un procedimiento sencillo para tratar con este fenómeno en los modelos de ganancia, cuya idea fundamental es ponderar la puntuación observada por la fiabilidad. Dicho autor, recomienda que cuando exista una correlación imperfecta entre dos puntuaciones, hacer un ajuste a la puntuación inicial, donde la puntuación inicial sea la puntuación inicial, más el producto de uno, menos la fiabilidad test-retest, por la media de la muestra total, menos el puntaje inicial (ver ecuación IV.21).

$$X' = x + (1 - r_{xx})\,(\mu - x) \qquad \textbf{Ec. IV.21}$$

Donde X′, es la puntuación inicial ajustada, x la puntuación inicial, r_{xx} la fiabilidad test-retest y μ la media para la muestra total. En definitiva, dicho autor da más peso a la puntuación real si la fiabilidad es alta, y si es baja le da más peso a la media global.

Aunque este procedimiento reduce el sesgo asociado con la regresión a la media, no lo elimina (Linn, 1980; Reisner, Alkin, Boruch, Linn & Millman ,1982; Roberts, 1980; Tallmadge 1982) además, este planteamiento es un problema en diseños con grupos no equivalentes o cuando se trabaja con grupos que presentan valores iniciales muy diferentes, y se pretende comparar las diferencias de ganancia entre dichos grupos.

Krause y Pinheiro (2007) proponen el uso de un enfoque de modelado y simulación en el que se compare un modelo "nulo" para producir conjuntos de datos simulados sin efectos de tratamiento, con el modelo original que se cree contaminado por el efecto de regresión a la media.

Rocconi y Ethington (2009) mantienen que un simple ajuste de la puntuación inicial ayudará a controlar el fenómeno de regresión y a proteger la validez interna. Siguiendo esta idea, Castro et al. (2009), quienes consideran que la existencia de una correlación negativa entre el estatus y la pendiente de crecimiento puede ser un posible indicador del ERM, utilizan para paliarlo, una puntuación inicial ajustada, empleando el estatus inicial del alumno como el principal predictor de la tasa de crecimiento.

Una posible alternativa cuando se desea trabajar con el residuo de las aulas para el estudio del VA de las mismas o la realización de ranking, consiste en llevar a cabo un ajuste de estos residuos asociados con el estatus y el crecimiento, mediante una regresión simple, tal y como plantea Navarro (2013).

Marsh y Hau (2002), tras simular distintos modelos multinivel concluyen que la mejor forma de paliar el ERM no consiste en realizar un ajuste del crecimiento previo sino utilizar dicha variable como covariable del modelo.

Un alternativa más eficiente y eficaz, acorde a lo estudiado sobre el ERM, consiste en cambiar el punto de partida utilizado. En esta línea, Rogosa (1995) estudió como el valor de la correlación y la intensidad está influido por la elección de un determinado punto de partida y, comprobó que la modificación del mismo elimina la correlación negativa entre el intercepto y la pendiente y, por lo tanto, el posible efecto de regresión a la media.

Navarro (2013) en su tesis, con el fin de comprobar la eficacia de algunas de estas estrategias, desarrolla diferentes modelos, en un diseño multinivel (tiempo, estudiante y escuela), con cuatro puntuaciones temporales. Siguiendo la idea de Rogosa (1995), plantea dos modelos donde modifica el punto de partida, en un primer modelo utiliza la segunda aplicación como estatus y en el segundo utiliza la tercera aplicación como punto de partida. También desarrolla un modelo en el que usa el predictor definido por Castro et al. (2009). Así mismo, prueba otro modelo donde utiliza el rendimiento previo como covariable, tal y como plantearon Marsh y Hau (2002), y finalmente realiza un análisis de regresión con los residuos del estatus y la pendiente de las escuelas para evidenciar si es una buena alternativa cuando se trabaja con los residuos de las escuelas. Los resultados muestran que los modelos que modifican el punto de partida, tal y como planteó Rogosa (1995), son los que proporcionan un mejor ajuste y consiguen eliminar la correlación negativa entre el intercepto y la pendiente (posible indicador del ERM). La aplicación de una regresión simple da unos resultados similares al de los dos modelos anteriores y genera residuos libres del ERM. El modelo que usa el predictor definido por Castro et al. (2009) elimina el ERM, pero también la varianza de crecimiento de los estudiantes, esto también ocurre en el modelo en el que se usa la primera aplicación como covariable (Marsh & Hau, 2002), sin embargo, dicho modelo no elimina el ERM.

Otra alternativa posible es estudiar si la puntuación observada es mayor o menor que la predicha por argumentos puramente estadísticos. En lugar de tener en cuenta la puntuación general de una escuela, se podría examinar las puntuaciones de cada escuela por curso; por ejemplo, los resultados de los estudiantes de cuarto curso en la escuela a la que asisten. Se trataría, en vez de comparar la puntuación media del estudiante con las puntuaciones medias de otros estudiantes de ese mismo curso en

todo el estado, o con las puntuaciones de los estudiantes de cuarto curso en otras escuelas con características demográficas similares, comparar sus puntuaciones de ese cuarto curso con sus puntuaciones anteriores. Si los estudiantes que estaban por encima o por debajo del promedio mejoran más de lo predicho por la regresión a la media, las escuelas están teniendo éxito

IV.1.5. Influencia de los estimadores bayesianos asociados al efecto de regresión a la media y la fiabilidad de las estimaciones, en las puntuaciones de los estudiantes a lo largo del tiempo

Otra posible causa del ERM, en un modelo longitudinal y que además puede afectar a la fiabilidad en las estimaciones de las puntuaciones de los estudiantes, son los estimadores bayesianos conocidos como BLUP, por sus siglas en inglés (*Best Linear Unbiased Predictor*).

Como bien indica Gelman, Hesketh, Long y Skrondal (2008) en un modelo multinivel, los efectos de grupo u_j (residuos del nivel superior, por ejemplo, los centros) son variables aleatorias que se supone que siguen una distribución normal. Su distribución se resume en dos parámetros: la media (se fija a cero) y la varianza de los centros σ_u^2. La varianza se estima junto con los otros parámetros del modelo: la varianza dentro del grupo (intrasujeto), varianza entre sujetos (σ_e^2) y la media global (β_0). Sin embargo, para hacer comparaciones entre grupos se puede querer obtener una estimación o valor predicho de u_j para cada grupo. Los u_j se calculan después de ajustar el modelo y se basan en las estimaciones de los parámetros del modelo ($\beta_0, \sigma_u^2, \sigma_e^2$) y los datos y_{ij}(puntuación del sujeto $i$en la escuela j).

En un modelo de un sólo nivel, se tiene un sólo conjunto de residuos. Para cada unidad muestral, el residuo se estima como la diferencia entre el valor observado de y_{ij} y el valor predicho por el modelo ajustado, $\hat{y}_{ij}$.En un modelo multinivel, el residuo total es $u_j + e_{ij}$ (residuos del estudiante) los cuales se estiman como :

$$r_{ij} = y_{ij} - \hat{y}_{ij} = y_{ij} - \hat{\beta}_0 \qquad \textbf{Ec. IV.22}$$

Donde r_{ij} es el residuo total, y_{ij} la puntuación del sujeto, $\hat{y}_{ij}$ el promedio de los sujetos y $\hat{\beta}_0$ el promedio global.

Es necesario dividir esto en estimaciones separadas de u_j y e_{ij}. Un punto de partida para una estimación de u_j podría ser tomar la media de $= y_{ij} - \hat{\beta}_0$ para el grupo j. Esto, se denomina a veces residuo medio bruto:

$$\bar{r}_j = \bar{y}_j - \hat{\beta}_0 \qquad \textbf{Ec. IV.23}$$

Donde $\bar{y}_j$ denota la media muestral de y_{ij} en el grupo j.

Para obtener una estimación del residuo para el grupo j, se multiplica el residuo bruto por un factor, llamado factor de suavización (k):

$$\hat{u}_j = k\bar{r}_j \qquad \textbf{Ec. IV.24}$$

Donde

$$\mathrm{k} = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + (\hat{\sigma}_u^2 / n_j)} \qquad \textbf{Ec. IV.25}$$

Los residuos estimados $\hat{u}_j$ se denominan residuos "*shrunk" o "shrunkage*" porque encogen o suavizan los resultados estimados hacia la media global de aquellas escuelas con poca fiabilidad en sus estimaciones. El factor de contracción o suavización k es siempre menor que o igual a 1, de modo que $\hat{u}_j$ (residuo del centro) será menor o igual que los residuos medios brutos $\bar{r}_j$. Para un n_j (tamaños de los centros) grande, el factor de contracción estará próximo a 1, y, por lo tanto, n_j estará próximo a $\bar{r}_j$. También habrá poca contracción (k cerca de 1) cuando $\hat{\sigma}_e^2$ es pequeño en relación con $\hat{\sigma}_u^2$. El factor de suavizado o contracción será notablemente menor que 1 cuando n_j es pequeño o, $\hat{\sigma}_e^2$ es grande en relación con $\hat{\sigma}_u^2$ (alta variabilidad dentro del grupo). En cualquier caso, cuando existe poca información sobre el grupo, la contracción del residuo bruto tiende hacia cero con el resultado de que la media del grupo $(\hat{\beta}_0 + \bar{r}_j)$ tiende hacia la media global $(\hat{\beta}_0)$. Estos residuos de contracción también se denominan estimaciones ponderadas con precisión, porque se tiene en cuenta la fiabilidad en su estimación. Las estimaciones poco fiables con un n_j pequeño se suavizan hacia la media global, mientras que las estimaciones fiables basadas en n_j grandes permanecerán cerca de su valor bruto.

De esta manera, mientras que los efectos fijos, generalmente se estiman mediante mínimos cuadrados generalizados, como un promedio calculado en base a todos los niveles del efecto en la población (Robinson, 1991). Los modelos

multinivel con coeficientes aleatorios estiman los residuos de las escuelas o los maestros con los estimadores bayesianos (BLUP), desarrollados por Henderson (1950) en experimentos de genética animal. Robinson (1991) apunta que, en la literatura, estas estimaciones son consideradas no sesgadas porque el valor esperado de los efectos aleatorios estimados sobre la población de los grupos es igual a la media de los efectos aleatorios verdaderos. Sin embargo, basándose en la definición actual de sesgo que refleja la preocupación natural de los interesados en la evaluación, las estimaciones BLUP están sesgadas, ya que, dado un verdadero efecto aleatorio para una escuela o maestro, hay una diferencia entre el valor esperado de las estimaciones BLUP de ese efecto y el valor verdadero.

Dicho predictor lineal insesgado, en los modelos lineales mixtos, minimizan el Error Cuadrático Medio (MSE, por sus siglas en inglés) de la predicción (Harville, 1990). El MSE está formado por dos componentes, uno referido a cualquier sesgo de estimación sistemática, y otro que refleja la variación aleatoria del valor esperado del estimador. Esta propiedad del MSE resulta de la introducción de un sesgo estadístico en la estimación, procedente de los estimadores bayesianos, para suprimir el componente aleatorio del MSE. Como resultado, la estimación BLUP se ve a menudo como una posible solución al problema de las estimaciones muy inestables de los efectos producidos en la situación de los profesores con clases muy pequeñas. Esto se afirma en la literatura de varias maneras. Por ejemplo, Goldstein (1997), al considerar la estimación de los efectos escolares, señala que las estimaciones BLUP son "conservadoras", en el sentido de que cuando hay poca información en una escuela (es decir, pocos estudiantes), las estimaciones están cercanas al promedio de todas las escuelas. Phillips y Adcock (1996) afirman que una solución multinivel ofrece estimaciones más estables en las escuelas más pequeñas.

Los estimadores BLUP contraen ("shrink") las estimaciones de los efectos de escuelas y profesores hacia la media general de las observaciones que dependen de la relación entre los componentes de la varianza involucrados, así, reducen la varianza de la estimación de los efectos de los centros en comparación con los modelos de efectos fijos (Navarro, 2013). De esta manera, las estimaciones BLUP también se pueden considerar una regresión hacia la media basada en los

componentes de varianza y covarianza asociada a los efectos aleatorios del modelo, que tiende a encoger o suavizar los resultados hacia la media global de aquellos centros con poca fiabilidad en sus estimaciones o que, cuentan con poca muestra de estudiante y han obtenido puntuaciones extremas en la primera medida (Sanders & Wright, 2008). Por tanto, si el tamaño muestral de un centro es pequeño, la precisión de su estimación será baja y los efectos reales de aquellos centros que puedan ser eficaces tenderán a no diferenciarse de la media y, por el contrario, centros con efectos reales bajos se parecerán a los resultados medios. Así, el encogimiento se verá afectado por la precisión de la media de la escuela y de la varianza entre centros (McCaffrey, 2003b). Una gran varianza supone poco encogimiento, mientras que, una varianza pequeña supone una mayor suavización de las estimaciones hacia la media general. Esto tiene la ventaja de que los valores estimados por encima o por debajo de la media regresan a la misma de acuerdo, a la magnitud de varianza respecto a la varianza total de las observaciones. Por lo tanto, las medias con valores extremos son atenuadas por la variabilidad subyacente y el riesgo de interpretar mal los datos se reduce (Little, Miliken, Stroup, Wolfinger & Schabenberger, 2007).

Si los datos de los centros son poco fiables (poca muestra) no se diferenciarán de la media general, por lo que no se tomarán decisiones equivocadas acorde a los resultados obtenidos. Sin embargo, una desventaja es que se desconoce si el efecto de los centros se debe a sus procesos educativos, ya que, aunque los estudiantes tengan resultados óptimos se parecerán a la media general.

El principal rasgo característico de las estimaciones BLUP es que utilizan los datos que proceden de otras escuelas, para estimar los efectos de un determinado centro, dando lugar al efecto de encogimiento ("*shrunk*"), donde los residuos medios estimados son ponderados por la fiabilidad (ver ecuación IV.26)[10], de tal manera, que tal y como se ha comentado en los párrafos anteriores, los centros que cuentan con tamaños muestrales muy pequeños tenderán a no diferenciarse de la media global y serán poco fiables (Armein, 2008; Gaviria & Castro, 2005).

$$\hat{\mu}_{0j} = \lambda_j(\bar{r}_j) \qquad \textbf{Ec. IV.26}$$

[10] En Gaviria y Castro (2005, p.71) se detalla la ecuación.

En la ecuación anterior IV.26 $\hat{\mu}_{0j}$ hace referencia a la media de los residuos de la escuela j (valor añadido); $\bar{r}_j$ es la media de los residuos de los sujetos en ese centro, que se encuentra multiplicado por λ_j (que es el factor de encogimiento, de ponderación). Dado que suele ser la fiabilidad del estimador, será menor o igual a 1, tendiendo a dicho valor según aumente el número de sujetos.

Dicha ponderación sobre las estimaciones de los centros se conoce como estimador empírico bayesiano:

$$Y_j + (1 - \lambda_j)\overline{Y} \qquad \textbf{Ec. IV.27}$$

Así, λ_j es la fiabilidad del rendimiento estimado en un centro (Y_j), es decir, la media de los resultados de los alumnos; $\overline{Y}$ es el promedio general. La puntuación observada de un centro tendrá un peso proporcional a la fiabilidad de la medida. Si la fiabilidad de dicha puntuación es alta, se añade a dicha puntuación un mayor valor, si es baja se valora más la media global (Navarro, 2013).

En la implementación de sistemas de rendición de cuentas ("*accountability systems"*) los efectos de los maestros o de los centros se estiman mediante el procedimiento BLUP, para ajustar las estimaciones de valor añadido de los mismos. En este sentido, dichas estimaciones juegan un papel muy importante, ya que en una estimación BLUP el efecto, por ejemplo, de un profesor individual, puede ser visto como una combinación de dos fuentes de información, la disponible para ese profesor específico y la información sobre todos los profesores que se están evaluando. Así, si el objetivo de un programa de rendición de cuentas es clasificar a todos los maestros de un distrito con respecto al logro promedio de los estudiantes en sus clases al final del año escolar, una clasificación basada en estimaciones BLUP, para cada maestro, combinaría la media de rendimiento estudiantil observada para ese profesor, con el promedio general de las medias estudiantiles para todos los maestros en el distrito. El peso asignado a la media observada en esta combinación dependería en parte de la cantidad de información disponible para el profesor individual. Si la media de un profesor individual estuviera basada en un gran número de estudiantes, el peso en la media observada sería grande y la estimación de "encogimiento" resultante no sería muy diferente de la media observada. Por el contrario, si el tamaño de la clase para un profesor fuera muy pequeño, la estimación

BLUP encogería la media observada hacia la gran media de todos los profesores (Tate, 2004).

Debido a que las estimaciones del valor añadido basadas en un menor número de estudiantes con logros difíciles de predecir serán menos precisas, el procedimiento podría conducir a la toma de una serie de decisiones inadecuadas, ya que, por ejemplo, en ausencia del "encogimiento" producido por las estimaciones BLUP, los profesores con muy pocos estudiantes podrían ser más propensos a terminar en los extremos de la distribución. Kane y Staiger (2002), encontraron que las escuelas pequeñas tenían más probabilidades de tener grandes cambios en el rendimiento a través de los años, que las escuelas grandes.

A pesar de las connotaciones que supone el uso de este tipo de estimaciones, hay poca evidencia de que los involucrados en los programas de rendición de cuentas aprecien plenamente las implicaciones de este sesgo diferencial, en la presencia de variación del tamaño de la escuela o de la clase; Bryk y Raudenbush (1992) representan una de las pocas excepciones a esta falta de atención.

En un estudio realizado por Tekwe et al. (2004) comparan modelos de efectos fijos y aleatorios estimados con BLUP, sin y con predictores, para estudiar el valor añadido de las escuelas. Los autores concluyeron que no había diferencias significativas entre los modelos BLUP sin y con predictores (correlación de Pearson entre 0,97 y 1). La correlación entre los modelos de efectos fijos y aleatorios (sin covariables) osciló entre 0,50 y 0,90.

En otro estudio, llevado a cabo por Herrmann, Walsh, Isenberg y Resch (2013), sobre cómo el procedimiento BLUP afecta a las estimaciones del valor añadido de los profesores, con estudiantes difíciles de predecir (bajo rendimiento previo), en el contexto de los sistemas de rendición de cuentas, encontraron que es más difícil predecir el logro de grupos particulares de estudiantes con bajo rendimiento previo que aquellos que tienen un rendimiento medio. Los maestros de estudiantes con bajo rendimiento previo tienden a tener estimaciones de valor añadido menos precisas; en estos casos los procedimientos BLUP aumentan la precisión de sus estimaciones, ya que reducen el valor absoluto de las estimaciones de valor añadido. Además, encontraron que las estimaciones BLUP, no tienen ningún

efecto estadísticamente significativo sobre la probabilidad relativa de que los maestros de alumnos con bajo rendimiento previo reciban consecuencias.

Posteriormente, en un estudio realizado por Guarino, Maxfield, Reckase, Thompson y Wooldridge (2015) aplicaron el método de estimación BLUP a un conjunto de datos simulados, donde se conocía el verdadero efecto del profesor. En dicha investigación se evalúa la capacidad de las estimaciones BLUP para clasificar adecuadamente a los maestros y eliminar el sesgo. Encontraron que estos estimadores funcionan bien bajo la asignación aleatoria de maestros a las aulas, y bajo la asignación no aleatoria, los estimadores controlan explícitamente el mecanismo de asignación proporcionando estimaciones fiables, ya que dichas estimaciones no cambian mucho las clasificaciones reales de los datos.

Los debates sobre las estimaciones BLUP en la literatura estadística, suelen dejar claro que la cantidad de "contracción" asociada al efecto estimado de cualquier escuela o maestro depende en parte, del número de alumnos en la escuela o clase.

Las estimaciones BLUP, a pesar de los posibles inconvenientes que ofrecen, como puede ser el ERM, proporcionan algunas ventajas como las siguientes:

- Protegen contra estimaciones falsas debido a la escasez de datos (Sanders, 2006).
- Minimizan la probabilidad de que una escuela sea clasificada erróneamente como muy buena o extremadamente mala por casualidad (Franco, 2006).
- Producen la máxima correlación entre la estimación y el efecto real (Searle et al., 1992).

IV.1.5.1. Otros fenómenos estadísticos asociados a la fiabilidad en las estimaciones de las puntuaciones de los estudiantes a lo largo del tiempo.

Otros fenómenos que pueden afectar a la fiabilidad de las estimaciones en las puntuaciones de los estudiantes a lo largo del tiempo, son el conocido efecto suelo y efecto techo, debido al diseño de las escalas verticales.

Estos fenómenos están presentes cuando un estudiante alcanza un valor extremo, ya sea un valor mínimo (suelo) o un valor máximo (techo), y, por lo tanto, sólo hay una dirección en la que las mediciones posteriores pueden cambiar, incrementándose o disminuyéndose.

Wang, Zhang, McArdle y Salthouse (2009) investigaron las influencias del efecto techo y del efecto suelo en el análisis de curvas de crecimiento con datos longitudinales y, exploraron posibles soluciones para tratar dichos fenómenos. Para ello, simularon una serie de datos y compararon las estimaciones de los parámetros de los modelos que no están contaminados por dichos efectos, con los verdaderos valores de los parámetros. Entre los métodos probados para la eliminación de dichos artefactos utilizaron, la eliminación por lista (Listwise Deleting) que consiste en eliminar aquellos casos que presentan el efecto suelo y el efecto techo, dicho método tiene la desventaja obvia de la pérdida de datos y la reducción del tamaño de la muestra. Otro método consistió en considerar a los datos afectados por dichos efectos como datos perdidos no debidos al azar[11]. Y finalmente utilizaron el Modelo Tobitárquico Jerárquico[12].

Tobin en el año1958, estableció dicho modelo:

Ec. IV.28

$$y = y^* \quad si\ y^* < c$$
$$y = c \quad si\ y^* \geq c$$

La variable y^* es la puntuación latente (no observable), libre de asumir todos los valores, sin verse afectada por el valor de c (límite o censura). La variable y es la puntuación observada, no puede exceder el punto de censura, c. Así, y^* se conoce cuando es menor que c. Cuando y^* es mayor o igual que c, sólo se conoce un límite inferior para y^*. La variable de interés es en realidad y^*, no y, pero desafortunadamente y^*, no es observable, de tal manera que y proporciona la mejor información disponible sobre y^*. Los maestros encuentran esta situación a diario. Por ejemplo, tres estudiantes que obtienen una puntuación del 100% en un examen de matemáticas; a medio plazo los estudiantes no son iguales en términos de su logro matemático subyacente (y^*), aunque reciben calificaciones idénticas en la prueba de matemáticas (y). Aquí, el punto de censura, c, representa la puntuación máxima alcanzable en la prueba, que es 100%. Si la prueba de matemáticas contiene elementos adicionales de dificultad suficientemente alta, se podría distinguir las

[11] Este tipo de patrón es conocido como MNAR, para más información consultar: Fitzmaurice, Laird y Ware., 2004; Little y Rubin,1987; Rubin, 1976

[12] Para más información consultar: Cowles, Carlin, y Connett, (1996), Hajivassiliou, (1994) y Matthew (2010)

actuaciones de estos tres estudiantes. La prueba de matemáticas simplemente tiene un efecto techo demasiado bajo, para discriminar entre los altos niveles de logro de los tres estudiantes (Matthew, 2010)

En términos de modelado de la curva de crecimiento, referido a datos con efecto techo en diseños longitudinales, el modelo Tobit jerárquico se puede desarrollar mediante la sustitución de la función de crecimiento de la siguiente ecuación (Wang et al., 2009) como:

Ec. IV.29

$$y_{it}^{*} = b_{0i} + A(t)b_{1i} + e_{it}$$

$$y_{it} = y_{it}^{*} \text{`}para\ y_{it}^{*} > c,$$

$$y_{it} = c, para\ y_{it}^{*} > c$$

Donde y_{it}^{*} es la puntuación verdadera del participante i en la ocasión t; y_{it} es la puntuación observada y c es el límite del efecto techo. Los resultados de la simulación mostraron, en primer lugar, que ignorar los efectos de techo y suelo, suponen sesgos en las estimaciones de los parámetros porque los datos se tratan incorrectamente, ya que, en los análisis de los mismos, se consideran puntuaciones verdaderas de la habilidad. En segundo lugar, cuando se suprimen los valores extremos del análisis de datos o se consideraron datos faltantes, no se utiliza la información de los datos extremos. Los modelos de curva de crecimiento Tobit permiten utilizar toda la información disponible al analizar los datos, lo que permite una buena recuperación de los valores verdaderos de los parámetros. Matthew (2010) también realizó un estudio en el que simuló el efecto techo y probó el modelo de Tobit y Matthew (2010) llegando a la misma conclusión que Wang et al. (2009) y es que, dicho modelo permite una buena recuperación de los valores reales.

Otro efecto que también puede influir en las estimaciones de las puntuaciones de los estudiantes y en el ERM es el efecto Mateo que equivale a una correlación creciente entre las aplicaciones.

Una relación positiva entre el punto de partida de los estudiantes y su tasa de crecimiento es conocido como «efecto abanico» ("fanspread effect"), ya que aquellos estudiantes que puntúan más alto que otros en un pre-tests, a lo largo del tiempo crecen más, en términos de puntuación absoluta y relativa, que aquellos que tienen

la misma experiencia educativa pero inicialmente habían puntuado más bajo (Cook & Campbell, 1979). Dicha relación positiva puede ser un indicador de lo que se conoce como Efecto Mateo[13] (Merton 1968, 1988), donde los sujetos con puntuaciones iniciales altas tenderán a crecer más. Un ejemplo, de dicho fenómeno se observó por primera vez en los resultados de la Encuesta Nacional de Igualdad de Oportunidades Educativas (National Assessment of Education Progress, NAEP, 1977) donde los grupos socioeconómicos y étnicos que obtuvieron en los primeros grados puntuaciones algo más altas que otros, obtuvieron puntuaciones mucho más altas en los grados posteriores; así la brecha o la ventaja acumulativa aumentó constantemente con el paso del tiempo (Coleman et al., 1966).

Para hablar de efecto Mateo deben darse dos condiciones. En primer lugar, que las diferencias entre sujetos con altas y bajas habilidades permanezcan estables (es decir, los "ricos" sigan siendo “ricos” mientras que los "pobres" sigan siendo “pobres”). En segundo lugar, las diferencias en competencias entre buenos y malos deben aumentar con el tiempo (es decir, los "ricos" se hacen más "ricos", mientras que los "pobres" se vuelven más "pobres") (Morgan, Farkas & Hibel, 2015). En este sentido, las técnicas de modelado de curvas de crecimiento (Goldstein, 1995; Raudenbush & Bryk 2002; Singer & Willett, 2003) tienen una serie de propiedades que son particularmente útiles en el estudio de dicho efecto, ya que los coeficientes calculados permiten estudiar los valores iniciales de la variable estudiada y, las trayectorias de crecimiento para los subgrupos de población definidos. Estos coeficientes iniciales estimados y las trayectorias de crecimiento revelan, qué grupos de sujetos se están convirtiendo más "ricos" y "pobres" con el paso del tiempo.

Walberg y Tsai (1983), fueron los primeros autores que encontraron evidencia del efecto Mateo en un contexto educativo. Descubrieron que el logro científico general entre los adultos jóvenes depende de antecedentes educativos

[13] La expresión “efecto Mateo” se toma de un pasaje bíblico (Mateo 25:29) que dice que el que tiene se le dará y tendrá en abundancia; pero al que no tiene incluso lo que tiene se le quitará. A partir de dicho escrito, el sociólogo Robert K. Merton (1968,1988) acuñó el término efecto Mateo para explicar las discrepancias en el reconocimiento recibido por eminentes científicos e investigadores desconocidos en un trabajo similar. el científico más conocido siempre recibe más reconocimiento que el poco conocido a pesar de realizar el mismo trabajo.

previos relevantes (incluidos: aspectos psicológicos de los entornos experimentados en las escuelas, socioeconómicos, características de las familias y los grupos étnicos), de la actividad educativa actual y de la motivación. Los antecedentes educativos predicen la actividad educativa actual y la motivación y, a su vez, los tres contribuyen al logro científico.

Más tarde el psicólogo canadiense Stanovich (1986) utilizó dicho término cuando descubrió que los menores con bajo progreso en el aprendizaje de la lectura y escritura durante las etapas más tempranas, se muestran más lentos en años posteriores, ya que sentían su fracaso y se resistían a pasar por otra mala experiencia y, por lo tanto, leían y escribían menos. Sin embargo, los estudiantes que conseguían destreza en la lectura y escritura en las primeras etapas educativas, reforzaban dichas aptitudes, lo que les permitía mejorar cada vez más en estas capacidades. De esta manera, parece que el efecto Mateo en la lectura y escritura predice que la brecha entre buenos y pobres lectoescritores aumenta con el paso del tiempo. A partir del estudio de Stanovich (1986), el efecto Mateo se ha estudiado casi de manera exclusiva en el campo del aprendizaje de la lectura.

Si bien ha habido estudios posteriores sobre el efecto Mateo vinculado con el aprendizaje lector (McNamara, Scissons & Gutknecth, 2011), hay que señalar que no todos los investigadores apoyan su existencia. Por ejemplo, Protopapas, Sideridis, Mouzaki, y Simos (2011), en su estudio sobre comprensión lectora, encontraron que la brecha, entre alumnos con altas puntuaciones y bajas, no se incrementaba con el paso del tiempo. Morgan, Farkas y Wu (2011) demostraron un efecto Mateo para las matemáticas, pero no para la lectura.

A la vista de lo comentado en los diferentes apartados del capítulo IV, se puede concluir que el ERM siempre esta presente entre la primera ocasión de medida y el resto de aplicaciones. Por lo tanto, una estimación precisa de los niveles de logro de los estudiantes, es poco probable en los modelos de ganancia que cuentan con dos ocasiones de medida, ya que el efecto de regresión a la media entre la primera y la segunda aplicación parece inevitable. En los diseños con más de dos mediciones donde se siguen a los mismos sujetos, el ERM no es un problema en la estimación de los resultados finales, ya que solo existe efecto de regresión a la media entre la primera ocasión de medida con el resto de aplicaciones, por lo tanto, a partir

de la segunda medición, ya se puede empezar a estimar el rendimiento de los estudiantes de manera precisa, pues, la correlación entre la segunda ocasión de medida con el resto de mediciones (a excepción de con la primera) no se ve afectada por dicho artefacto y, lo mismo para las demás aplicaciones. Sin embargo, se desconoce el número de mediciones necesarias para conocer de manera fiable el rendimiento del estudiante.

Aunque el comportamiento del ERM depende del patrón de correlación existente entre el intercepto y la pendiente, hay una serie de estrategias que permiten modificar el sentido y la direccionalidad de dicha relación, entre ellas destaca la modificación del punto de partida tal y como recomienda Rogosa (1995), que permite eliminar un posible indicador del ERM como es la relación negativa entre el estatus y la pendiente en los modelos con más de dos mediciones (Navarro, 2013)

Además del ERM existen otros artefactos estadísticos como son el efecto suelo, el efecto techo o el efecto Mateo que también deben ser tenidos en consideración dentro de los diseños longitudinales, ya que pueden afectar a la fiabilidad con la que se estiman las puntuaciones de los estudiantes a lo largo del tiempo.

PARTE METODOLÓGICA

CAPÍTULO V: DISEÑO DE UNA SIMULACIÓN MONTE CARLO BASADA EN MODELOS MULTINIVEL DE CARÁCTER LONGITUDINAL

Tal y como se ha descrito en el capítulo anterior (Capítulo IV), las características de las evaluaciones longitudinales y el número de mediciones, pueden afectar a la fiabilidad con la que se estiman las puntuaciones de los estudiantes. A nivel técnico, que una evaluación basada en tres mediciones (número mínimo de ocasiones de medida requeridas para el desarrollo de la curva de crecimiento de los estudiantes) proporcione estimaciones del rendimiento lo suficientemente fiables para la toma de decisiones, supone un gran reto, ya que, no existen estudios que demuestren el número de aplicaciones necesarias para poder realizar un acercamiento al rendimiento real del estudiante, teniendo en cuenta ciertos artefactos estadísticos.

Según Furby (1973) el ERM es uno de los fenómenos más presentes en todos los diseños longitudinales. Así, posteriormente a las explicaciones dadas por dicho autor sobre el Efecto de Regresión a la Media, Nesselroade et al. (1980) realizaron un estudio en profundidad sobre el mismo, analizando mediante modelos jerárquicos lineales de dos niveles (siendo el primer nivel el tiempo y el segundo nivel los sujetos), los distintos patrones de correlación que se pueden encontrar entre la primera ocasión de medida y el resto de aplicaciones, así como las implicaciones que dichos patrones tienen sobre el ERM.

Sin embargo, ni estos autores, ni otros muchos, demuestran empíricamente el patrón de correlación que se genera entre las puntuaciones de los estudiantes en las distintas ocasiones de medida, más allá de la primera medición. En esta tesis, se estudiará dicho patrón, lo que permitirá conocer el número de aplicaciones necesarias para obtener estimaciones fiables de los niveles de logro de los estudiantes.

También el tamaño muestral es un aspecto muy importante en las evaluaciones educativas, que utilizan datos longitudinales con carácter multinivel, ya que, generalmente en estos diseños anidados, se requieren al menos de 20 sujetos por unidad del nivel macro (Goldstein, 1997, Bryk & Raudenbush, 2004; Raudenbush & Bryk, 1986; Snijders & Bosker 2012), con el fin de obtener

estimaciones óptimas, con pocos errores de medida. Sin embargo, en las evaluaciones educativas longitudinales, no siempre es posible contar con un gran número de aulas o centros educativos compuestos por un mínimo de 20 estudiantes, debido, entre otros factores, a la mortandad experimental. Lo que puede conllevar a estimaciones sesgadas (Armein, 2008; Goldstein, 1997; Lee & Ban, 2010; McCaffrey et al., 2004; Phillips & Adcock, 1996; Tong & Kolen, 2007), puesto que la pendiente de crecimiento de dichas aulas o centros, tenderá a no diferenciarse de la media global debido a las estimaciones BLUP.

En este sentido, las investigaciones educativas basadas en el modelaje multinivel, no han prestado mucha atención a como el tamaño del nivel macro (generalmente aulas o escuelas) puede afectar en la estimación de la pendiente de crecimiento de dicho nivel. Muestra de ello es que en educación una quinta parte de los estudios basados en modelos multinivel, tienen grupos de menos de 30 unidades (Dedrick et al., 2009). Esta falta de cuidado en lo que respecta al tamaño muestral, puede ser debido a la escasez de investigaciones que estudian el efecto del mismo en el campo educativo, por ello en dicho trabajo de investigación, también, se analizará de forma más general, si el número de estudiantes por aula afecta en la estimación de su pendiente de crecimiento.

Partiendo de estas ideas, la finalidad principal de dicha tesis es, a partir de la realización de una simulación Monte-Carlo, estudiar cómo influyen algunos factores propios de las evaluaciones educativas basadas en datos longitudinales y anidados, en las estimaciones de los niveles de logro de los estudiantes a lo largo del tiempo, y, por ende, en la pendiente de crecimiento de las aulas.

Este capítulo, se centra principalmente en tres grandes apartados, la presentación de los problemas de investigación, los objetivos vinculados a cada uno de ellos, y la metodología empleada, dentro de la cual se especifica: el diseño del experimento (ya que se trata de una simulación), la descripción de las variables y el plan de análisis referido: 1) al procedimiento de generación de los datos, 2) la precisión en la recuperación de parámetros y 3) el procedimiento de generación de modelos.

V.1. Presentación de los problemas de investigación

El efecto de regresión a la media se encuentra presente en los diseños longitudinales donde se sigue a los mismos estudiantes, entre la primera ocasión de medida y el resto de mediciones, pero ya desaparece entre la segunda y la tercera aplicación (Nesselroade et al., 1980). Por lo tanto, los modelos de ganancia caracterizados porque solo cuentan con dos mediciones, se encuentran afectados por este artefacto estadístico, lo que puede influir en la fiabilidad de las estimaciones proporcionadas por dichos modelos.

El comportamiento del efecto de regresión a la media entre la primera ocasión de medida y el resto de aplicaciones depende del patrón de correlación encontrado entre las puntuaciones de los sujetos. Si este patrón se mantiene estable, es que el ERM entre la primera ocasión de medida y el resto de ocasiones es constante. El ERM se intensifica a medida que dicho patrón va decreciendo (patrón simplex o autorregresivo), y puede mitigarse en el caso de que la correlación entre la primera ocasión y el resto de ocasiones sea creciente, pudiendo llegar a desaparecer dicho fenómeno cuando la correlación entre la primera medida y alguna de las mediciones realizadas es perfecta (valor de correlación igual a 1). Cuando el patrón de correlación es creciente se conoce como egresión a la media (Nesselroade et al., 1980).

Pero, ¿qué patrón se obtiene entre el resto de ocasiones de medida, más allá de la primera medición? y de acuerdo a ese patrón, ¿cómo de fiables son las estimaciones de los niveles de logro de los estudiantes en el modelo analizado?, y lo que es más importante, ¿qué condiciones generan ese patrón?, y, por tanto, ¿de que depende la fiabilidad con la que se estima el rendimiento de los estudiantes? Estas preguntas serán resueltas en este trabajo de investigación, ya que no se han encontrado respuestas basadas en evidencias metodológicas.

Tradicionalmente, tal como advirtió Blanco, González y Ordóñez (2009), y Nesselroade et al. (1980) se ha considerado que en los diseños longitudinales el patrón de correlaciones más común, entre las distintas ocasiones de medida, es el simplex o autorregresivo (Balzarini, Macchiavelli & Casanoves, 2005; Garson, 2013; Guttman, 1954; Humphreys, 1960, 1968; Kenny, 1979; Lunneborg & Lunnerborg,

1970; Pardo & Ruiz, 2013; Raudenbush. & Bryk, 1986; 2002; Ronald et al., 2010; Vallejo 1989; Vallejo & Fernandez, 1990).

Concretamente, desde la introducción al modelo simplex realizada por Guttman en el año 1954, la creencia de que el patrón esperable en este tipo de modelos era el decreciente se hizo muy presente cuando se trabajaba con datos anidados de carácter longitudinal, y algunos autores como, por ejemplo, Humphreys, (1960); Joreskog, (1970), Joeskog y Sorborm, (1977) y Wert y Hilton, (1977), han transmitido esta idea hasta hoy en día, sin que existan demostraciones empíricas que realmente lo demuestren.

Así, bajo esta concepción predominante de la existencia de un patrón decreciente en los diseños longitudinales, en una evaluación comprendida por cuatro ocasiones de medida, se espera que la correlación entre la primera ocasión y la segunda ocasión, sea superior a la correlación entre la primera ocasión y la tercera ocasión, correlación, que se espera que a su vez sea superior a la que se registra entre la primera ocasión y la cuarta medida.

Desde esta perspectiva y dentro del marco de las evaluaciones educativas basadas en datos anidados y longitudinales, el primer problema de investigación general de dicho capítulo es: **¿cómo afecta el porcentaje de dispersión fijado en los coeficientes aleatorios de los distintos niveles de anidamiento, así como la correlación entre los parámetros del intercepto y la pendiente, en la fiabilidad de las estimaciones de las puntuaciones de los estudiantes, y en el comportamiento del ERM?**

Para dar respuesta a dicho problema de investigación se simulan diferentes modelos, con una estructura propia de los modelos jerárquicos lineales, (tres niveles de anidamiento: tiempo, estudiante y aula) y longitudinales (nueve ocasiones de medida), en los cuales se manipulan diferentes condiciones experimentales.

La manipulación del porcentaje de varianza fijado en los diferentes coeficientes aleatorios de los distintos niveles (tiempo, estudiante y aula), así como el valor de correlación fijado entre el intercepto y la pendiente, permitirá estudiar el patrón de correlación obtenido entre las distintas ocasiones de medida, bajo las diferentes condiciones simuladas, y, por ende, el número de mediciones requeridas para obtener estimaciones fiables de las puntuaciones de los estudiantes. Asimismo,

mediante el análisis del patrón de correlación existente entre la primera medición y el resto de ocasiones, se podrá estudiar el comportamiento del ERM en los diferentes modelos simulados, lo que puede ser interesante con el fin de clarificar y completar el estudio de dicho fenómeno, explicado teóricamente en el Capítulo IV.

Como se ha comentado al inicio de este capítulo, también se presta atención, aunque con un carácter más general, al tamaño muestral necesario cuando se trabaja con datos anidados, ya que, un aspecto importante dentro del modelado multinivel es conocer que tamaño muestral es suficiente para una estimación precisa de la pendiente de crecimiento, pues en el contexto de una evaluación educativa real, generalmente no se trabaja con un alto porcentaje de centros educativos compuestos por más de 20 estudiantes por unidad macro (Dedrick et al., 2009). En este sentido, el segundo problema de investigación es: **¿Cómo afecta el tamaño de las aulas en la estimación de su pendiente de crecimiento?**

Con el fin de dar respuesta a dicho problema de investigación, también se realiza una simulación Monte Carlo de carácter longitudinal y multinivel donde se manipula el porcentaje de aulas con un tamaño determinado. Pues en las evaluaciones educativas que trabajan con datos anidados, una restricción fundamental es a menudo el tamaño de la muestra del nivel de anidamiento superior (Maas & Hox, 2005), ya que en los diseños longitudinales es muy común la pérdida de muestra, debido a que es posible que a lo largo del tiempo muchos estudiantes cambien de escuela, no acudan el día de la prueba, dejen la prueba en blanco, etc.; originando que en determinadas ocasiones no se supere los 20 casos por unidad. Cantidad minima recomendada para realizar un análisis multinivel (Bryk & Raudenbush, 2002; Goldstein, 1997; Raudenbush & Bryk, 1986), puesto que tamaños inferiores a 20 sujetos pueden ser un problema, ya que, los centros no tenderán a diferenciarse de la media global (Lockwood, Louis & McCaffrey, 2003) debido a las estimaciones BLUP. Incluso hay autores como Kreft y De Leeuw (1998) que recomiendan para realizar el análisis multinivel en condiciones óptimas, al menos 30 casos.

Snijders y Bosker (1999) desaconsejan el modelo multinivel si el número de conglomerados es inferior a 10, con pocas unidades en cada uno. Sin embargo, en los entornos aplicados, las exigencias de estas recomendaciones no siempre se

cumplen, lo que lleva a resultados potencialmente parciales. Por ejemplo, en una revisión de Dedrick et al. (2009), de los 99 estudios multinivel revisados entre el año 1999 y 2003, en 13 revistas de educación, psicología y sociología, el 21% tenía tamaños de muestra que no cumplían con la recomendación de un mínimo de 30 sujetos por unidad del nivel macro. Este hallazgo sugiere que los investigadores pueden no estar al tanto de las recomendaciones del tamaño de muestra necesaria para la estimación multinivel, o no tener los recursos para obtener muestras grandes, ya que reclutar grandes cantidades de estudiantes puede ser extremadamente difícil.

Aunque los modelos con tamaños de muestra pequeños pueden converger a una solución, y producir estimaciones estables de los parámetros, por todo lo comentado anteriormente, parece que el tamaño muestral puede producir sesgos en las estimaciones, afectando a las inferencias en la investigación aplicada. Por ello, la respuesta a este segundo problema de investigación, es importante, ya que, permitirá a los investigadores conocer y reconocer el efecto del tamaño muestral, en la fiabilidad de las estimaciones de la pendiente de crecimiento de las aulas o centros educativos.

Se considera que el presente estudio basado en dichos problemas de investigación, aportará a los especialistas e interesados en los sistemas educativos, especialmente en evaluaciones educativas de calidad, así como a las personalidades encargadas de las políticas educativas, conocer qué factores de las evaluaciones educativas, afectan y cómo afectan a la fiabilidad con la que se estiman las puntuaciones de los estudiantes a lo largo del tiempo. Lo que les permitirá tener una idea aproximada del número de medidas necesarias para poder realizar un acercamiento al rendimiento real de los estudiantes. Asimismo, podrán adquirir un conocimiento más claro del comportamiento del ERM en las evaluaciones longitudinales donde se siguen a los mismos estudiantes, y de cómo puede afectar el tamaño muestral en la estimación del crecimiento medio de los centros educativos, cuando se trabaja con datos anidados. Con todo ello, se pretende proporcionar a la comunidad educativa una serie de pautas respaldadas por la investigación, que sirva para la mejora de la calidad de las evaluaciones educativas y del sistema educativo en general.

V.2. Objetivos de investigación

En este capítulo de investigación, se pretende estudiar cómo la manipulación de las varianzas y la correlación entre el intercepto y la pendiente en datos anidados de carácter longitudinal, afectan en la precisión con la que se estiman los niveles de logro de los estudiantes a lo largo del tiempo y en el comportamiento del ERM (primer problema de investigación), así como el efecto del tamaño muestral de las aulas en la estimación de su crecimiento (segundo problema de investigación). Para dar respuesta al primer problema se analizan los patrones de correlación obtenidos en las puntuaciones de los estudiantes a lo largo del tiempo, y en el segundo el residuo de la pendiente del nivel de aulas.

Concretamente, el objetivo general del primer problema de investigación es **simular una serie de modelos longitudinales y multinivel (modelos jerárquicos lineales) que permitan estudiar el efecto de los porcentajes de varianza, así como de los valores de correlación fijados entre el intercepto y la pendiente, sobre los patrones de correlación generados entre las distintas ocasiones de medida.**

Para ello, como objetivos más específicos, dentro de este primer problema de investigación, se plantean los siguientes:

1. Garantizar una adecuada recuperación de las condiciones experimentales fijadas, antes de la generación definitiva de los resultados, a través de una simulación Monte Carlo de un modelo jerárquico lineal con tres niveles de anidamiento (tiempo, estudiante y aula) y nueve ocasiones de medida.
2. Estudiar el efecto de los factores manipulados en los distintos modelos simulados sobre la fiabilidad de las estimaciones de las puntuaciones de los estudiantes, y sobre el comportamiento del ERM, bajo tres situaciones distintas:
   a) Cuando solo se fija varianza en algunos coeficientes aleatorios del modelo.
   b) Cuando además de fijar variabilidad en el error intrasujeto, en el intercepto y en la pendiente, del nivel de aula o de estudiantes, se fija también, distintos valores de correlación entre el estatus y la pendiente de ese mismo nivel (uno de los dos niveles).

c) Cuando además de fijar variabilidad en todos los coeficientes aleatorios de los tres niveles (tiempo, estudiante y aula), se fija también distintos valores de correlación entre el intercepto y la pendiente del nivel de aula y del nivel de estudiante, de manera simultánea.

En cuanto al objetivo general del segundo problema de investigación, a partir de los coeficientes de varianza y los valores de correlación fijados en uno de los modelos desarrollados dentro del primer problema de investigación, se quiere **estudiar el efecto del tamaño de las aulas en la estimación de la pendiente de crecimiento de las mismas.**

Este objetivo general se concreta en los siguientes:

1) Garantizar una adecuada recuperación de las condiciones experimentales, antes de la generación definitiva de los resultados, mediante la simulación de distintos porcentajes de tamaños de aula (pequeños, medianos y grandes), manteniendo los valores de correlación y de varianza de uno de los modelos simulados en el primer problema de investigación.
2) Estudiar el efecto del tamaño de las aulas sobre la precisión con la que se estima la pendiente de crecimiento de las mismas, en tres circunstancias distintas:
   a) Cuando de la muestra total de aulas la mayoría son pequeñas (comprendidas por un mínimo de 5 estudiantes y un máximo de 10 estudiantes) y existe un porcentaje muy similar de aulas medianas (comprendidas por un mínimo de 11 estudiantes y un máximo de 19 estudiantes) y aulas grandes (comprendidas por un mínimo de 20 estudiantes y un máximo de 30 estudiantes).
   b) Cuando la muestra total de aulas esta compuesta, en su mayoría por aulas medianas (comprendidas por un mínimo de 11 estudiantes y un máximo de 19), y el resto de aulas son pequeñas y grandes, en proporciones similares.
   c) Cuando de la muestra total de aulas la mayoría son aulas grandes (comprendidas por un mínimo de 20 estudiantes y un máximo de 30), y el porcentaje restante esta formado de manera equitativa por aulas pequeñas y grandes.

V.3. Metodología

Como bien se ha venido adelantando en los apartados anteriores, el principal rasgo que caracteriza la metodología de este trabajo de investigación, es la realización de una simulación Monte-Carlo de caracter multinivel y longitudinal, con el fin de generar una serie de condiciones experimentales difíciles de obtener y manipular dentro del contexto real de las evaluaciones educativas.

En este sentido, en dicho apartado (*V.3 Metodología*) se describe en primer lugar el diseño del experimento, posteriormente se procede a una descripción detallada de cada una de las condiciones experimentales, así como de las variables dependientes de cada uno de los problemas de investigación. Finalmente, se concluye la metodología con el plan de análisis de datos donde se especifican los procedimientos seguidos para la generación de datos, los procedimientos utilizados para garantizar una adecuada recuperación de los parámetros fijados y la descripción de los modelos finales desarrollados.

V.3.1. Diseño del experimento

El análisis de los patrones de correlación entre las puntuaciones de los estudiantes en las diferentes ocasiones de medida no es tarea fácil. En principio puede parecer que con fijar distintos valores en los coeficientes aleatorios de los diferentes niveles (tiempo, estudiante y aula) y evaluando su grado de ajuste a la realidad que se analiza, se tendrían los datos necesarios para el estudio de la fiabilidad de las estimaciones. Sin embargo, el empleo de las diferentes herramientas metodológicas utilizadas, tiene como objetivo conseguir una visión completa y profunda sobre una realidad educativa que, al ser poco conocida y estudiada, ha dado lugar a la perseverancia de una serie de falacias metodológicas que serán cuestionadas en este estudio.

La utilización de una simulación experimental tiene la ventaja que permite determinar a priori la realidad objeto de estudio. De ahí que dicha técnica sea un procedimiento adecuado para crear un fenómeno educativo, pudiendo no sólo evaluar sus resultados sino también analizar el grado de ajuste de las estimaciones con referencia a una realidad que se conoce. Así, la simulación facilita la realización de una reflexión sobre las ventajas, inconvenientes, aportaciones metodológicas,

adecuación, exactitud, etc., de diferentes fenómenos que a veces en la realidad educativa es difícil encontrar, como por ejemplo la existencia de evaluaciones educativas con más de cinco mediciones realizadas a los mismos estudiantes, la posibilidad de fijar diferentes porcentajes de varianza en los distintos parámetros aleatorios del modelo, así como diferentes valores de correlación entre el intercepto y la pendiente, o bien la posibilidad de especificar el tamaño de las aulas. De esta forma, se justifica el empleo del uso de un estudio de simulación, ya que no es plausible obtener dicha información por otros procedimientos (Psychometric Society, 1979).

El esquema de desarrollo de un estudio de simulación experimental es análogo a las fases de la investigación clásica, ya que no es más que un experimento con algunas características particulares (Gaviria, 1989; Harwell, Stone, Hsu & Kirisci, 1996). Es necesario decir que dentro del experimento Monte-Carlo caben todos los diseños propios del contexto clásico.

Tal y como apunta Castro (1997) las simulaciones Monte-Carlo son aquellas que por cada condición experimental hay más de una muestra, por lo que da cabida al error muestral, constituyendo estudios que son equivalentes a los estudios experimentales clásicos, pero en vez de ser la unidad experimental el sujeto, es la muestra. Así, este es un estudio experimental clásico, donde se observa la influencia de las variables independientes manipuladas sobre la variable dependiente, objeto de interés. De forma coherente con el objetivo de simulación, los resultados se obtienen cuando se realizan varias réplicas de las mismas condiciones a una muestra cada vez distinta.

Para garantizar la validez interna y externa en los estudios clásicos, es condición necesaria la selección y asignación aleatoria de los sujetos a grupos, así como de los tratamientos. Sin embargo, en las simulaciones no hay extracción de muestras, sino generación pseudo-aleatoria de los datos, por lo que es necesario determinar un procedimiento para generar valores que formen una distribución aleatoria de acuerdo a unos parámetros determinados. Así, la validez interna depende de un procedimiento adecuado de generación de datos, mientras que la validez externa depende de las condiciones experimentales definidas por el

investigador. Normalmente estos procedimientos producen resultados altamente generalizables porque no hay incertidumbre (Harwell et al., 1996).

En este estudio de simulación Monte Carlo de un modelo multinivel, tal y como aparece en la siguiente ecuación:

$$Y_{tij} = \beta_{000} + \beta_{100}(T) + u_{00j} + u_{10j}(T) + r_{0ij} + r_{1ij}(T) + e_{tij} \qquad \textbf{Ec. V.1}$$

El primer nivel hace referencia al tiempo, dentro del cual se manipula la varianza del error intrasujeto (e_{tij}). Aunque el mínimo de mediciones requeridas para el desarrollo de modelos de crecimiento es tres, y son generalmente las que se encuentran en la mayoría de las evaluaciones longitudinales, por el coste que suponen. En los modelos desarrollados en dicha investigación se generan nueve ocasiones de medida considerando que cada una de ellas se realiza al final de un curso académico. Se ha concebido un alto número de mediciones, ya que existen algunas organizaciones como la *Northwest Evaluation Association* (NWEA) que llevan a cabo evaluaciones en las que realizan más de tres mediciones.

Concretamente, la NWEA evalúa los conocimientos científicos de los estudiantes comprendidos entre el tercer y noveno grado. Para ello, evalúa a cada estudiante dentro de su grado correspondiente realizándole cinco mediciones, tal y como aparece en la siguiente figura.

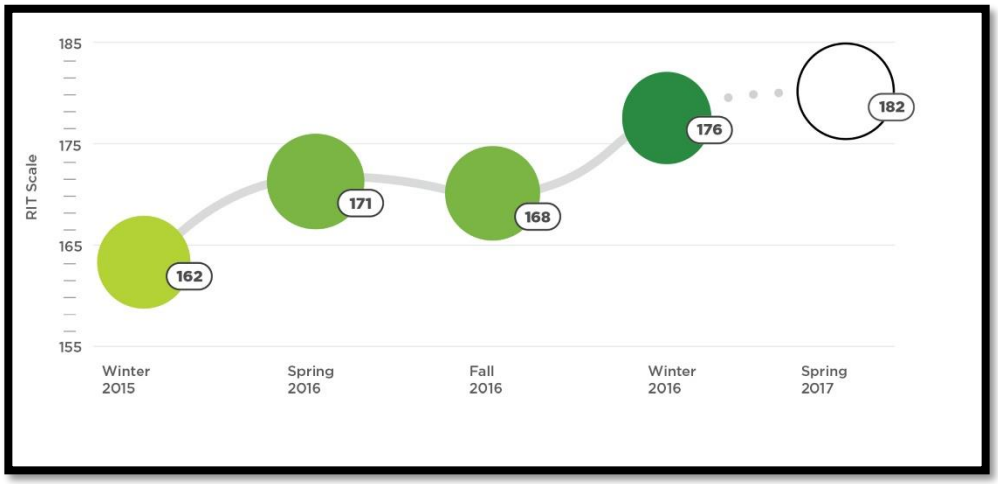


Figura V.4. Evaluaciones de la Northwest Evaluation Association (NWEA)
Fuente: https://www.nwea.org/map-growth/

El propósito de generar nueve ocasiones de medida es comprobar, bajo las distintas condiciones simuladas, cuál sería el número de mediciones necesarias para la obtención de estimaciones fiables del rendimiento académico de los estudiantes, y de esta manera comprobar, si el coste que supone realizar evaluaciones como las de NWEA compuesta por más de cuatro medidas, está justificado debido a la información que aportan, o bien, por el contrario, tres o cuatro medidas son más que suficientes para obtener información fiable sobre los niveles de logro de los estudiantes

El segundo nivel hace referencia a los estudiantes, dentro de este nivel se han manipulado cuatro características: el tamaño de la muestra ($N_{estudiantes}$)., el porcentaje fijado en la varianza del estatus ($\sigma(r_{0ij})$), el porcentaje fijado en la varianza de la pendiente ($\sigma(r_{1ij})$), y la correlación entre el intercepto y la pendiente de dicho nivel ($R_{estudiantes}$).

Finalmente, el tercer nivel se refiere a las aulas donde las características manipuladas son: el porcentaje fijado en la varianza del estatus ($\sigma(u_{00j})$), el porcentaje fijado en la varianza de la pendiente ($\sigma(u_{10j})$) y la correlación entre el intercepto y la pendiente de dicho nivel (R_{aulas}).

Asimismo, también se fijaron valores en el rendimiento promedio en la primera aplicación (β_{000}) y en la pendiente de crecimiento (β_{100}), aunque estos valores se mantienen constantes a lo largo de todos los modelos simulados, ya que no afectan al patrón de correlación originado entre las puntuaciones de los estudiantes, en las diferentes ocasiones de medida.

Debido al gran número de combinaciones que se pueden realizar con las diferentes características manipuladas, este estudio no tiene pretensiones de exhaustividad para todas las posibles situaciones que se pueden encontrar en las evaluaciones que trabajan con datos multinivel y longitudinales, ya que el número de combinaciones sería muy amplio. Por ello se va a trabajar bajo los determinantes que aportan información más relevante a los estudios del campo de la educación, con la finalidad de conseguir evaluaciones educativas de calidad.

V.3.2. Descripción de variables

Las definiciones de las variables independientes en la investigación experimental son los parámetros o características manipuladas. En la simulación Monte-Carlo estas variables independientes son conocidas como condiciones experimentales. Dichas condiciones son las características que poseerá cada muestra sobre las que se probará el modelo a evaluar.

De forma paralela a las variables independientes hay que definir las variables dependientes, donde se refleja el efecto de las distintas condiciones experimentales que suelen englobar comparaciones entre los valores obtenidos y los paramétricos, que en este caso son los valores conocidos, ya que son los empleados para generar los distintos modelos multinivel de carácter longitudinal.

V.3.2.1 Determinación de las condiciones experimentales

Las condiciones experimentales que se han seleccionado para realizar esta investigación, hacen referencia a factores propios de las evaluaciones educativas que utilizan datos multinivel y longitudinales.

En cuanto al **número de aulas**($\boldsymbol{N_{aulas}}$) aunque no se especifica un número concreto, si se fija una semilla para que siempre se simule el mismo número. Al no definir un número concreto de aulas, cada vez que se simula un modelo, el tamaño muestral del mismo puede variar, por ello, con la finalidad de mantener un tamaño de aula único y poder replicar los resultados se establece una semilla, que permite obtener siempre la misma secuencia de números aleatorios.

De esta manera, en el primer problema de investigación el número total de aulas que se generan son 400. En el segundo problema de investigación depende de la situación. En la *situación a,* caracterizada porque se establece que un 75% de aulas sean pequeñas (comprendidas por un mínimo de 5 estudiantes y un máximo de 10), un 12,5% de aulas medianas (comprendidas por un mínimo de 11 estudiantes y un máximo de 19) y el otro 12,5% de las aulas grandes (comprendidas por un mínimo de 20 estudiantes y un máximo de 30), se simulan un total de 942 aulas. En la *situación b* en la cual un 75% de las aulas se consideran medianas, un 12,5% de las aulas pequeñas y el otro 12,5% grandes, se simulan 654 aulas, y en la *situación c,*

caracterizada porque un 75% de las aulas son grandes, un 12,5% son pequeñas y alrededor de otro 12,5% son aulas medianas se simula un total de 464 aulas.

El **número total de estudiantes** ($\boldsymbol{N_{estudiantes}}$) se fija a 10.000 para todos los modelos simulados en el primer problema de investigación, así como para el segundo. Se fijaron 10.000 estudiantes, ya que se recomienda un minimo de 5.000 estudiantes para el estudio del crecimiento (Castellano & Ho, 2013) debido al gran número de parámetros a estimar.

En los modelos simulados tanto en el primer problema de investigación como en el segundo, se fija un máximo de 30 estudiantes por aula, ya que, según la LOMCE (2016) el número máximo de alumnos por aula en Educación Secundaria es de 30 estudiantes, aunque se contempla una subida de hasta un 10% de la ratio máxima establecida para hacer frente a las necesidades inmediatas de escolarización, de los estudiantes de incorporación tardía (LOMCE, artículo 87.2)

En cuanto al número mínimo de estudiantes por aula, aunque en Educación Secundaria Obligatoria es de 15 alumnos (Consejería de Educación, Juventud y Deporte, 2017), en el primer problema de investigación se fija un mínimo de 20 estudiantes por aula, por dos motivos. El primero de ellos, es que es poco común encontrar aulas con menos de 20 estudiantes, ya que en España la media de estudiante por aula en secundaria es de 25 alumnos (OECD, 2016a) y, por otro lado, a nivel metodológico se recomienda en el modelado multinivel que las unidades de análisis tengan un mínimo de 20 sujetos (Goldstein, 1997; Bryk & Raudenbush, 2002; Raudenbush & Bryk, 1986). Teniendo en cuenta el mínimo de unidades de análisis requerido, en el segundo problema de investigación el número de estudiantes por aula depende del tipo de aula. Las aulas denominadas como pequeñas están compuestas entre un mínimo de 5 estudiantes y un máximo de 10, las aulas medianas entre 11 y 19 estudiantes y las aulas grandes entre 20 y 30 estudiantes.

En cuanto al porcentaje de varianza fijado en la **pendiente del nivel de estudiantes** ($\boldsymbol{\sigma}(\mathbf{r_{1ij}})$), en los distintos modelos simulados en el primer problema de investigación, se ha fijado a: 0%, 1%,1,25%, 2,5%, 5%, 6,49%, 8,42%, 10%, 12,5%, 17,5%, 23, 75%, 25%, 33,3%, 35%, 40%, 47,5%, 50%, 65%, 95% y 100%, de la varianza total del modelo. En el segundo problema de investigación se fija a un 25%.

La varianza de la **pendiente del nivel de aula** $(\boldsymbol{\sigma}(\mathbf{u_{10j}}))$ del primer problema de investigación se ha fijado a: 0%, 1,25%, 5%, 10%, 12,5%, 12,27%, 23%, 75%, 25%, 33,3%, 47,5% y 95%, y en el segundo problema de investigación a un 5%.

Respecto a la varianza del **estatus del nivel de estudiante** $(\boldsymbol{\sigma}(\mathbf{r_{0ij}}))$**,** los valores se han fijado, en el primer problema de investigación a un 0%, 1,25%, 2,5%, 5%, 10%, 12,5%, 15%, 17,5%, 23, 75%, 25%, 33,3%, 35%, 37,7%, 40%, 47,5%, 48,96%, 50%, 65%, 85%, 95% y 100%, mientras que en el segundo problema se ha fijado a un 35%.

En la **varianza del intercepto del nivel de aulas** $(\boldsymbol{\sigma}(\mathbf{u_{00j}}))$**,** del primer problema de investigación, se han fijado los siguientes porcentajes: 0%, 1,25%, 2,5%, 3,36%, 4,35%, 5%, 12,5%, 15%, 23, 75%, 33,3%, 40%, 55% y 95%, y en el segundo problema de investigación, la varianza del intercepto del nivel de aulas se ha fijado a un 15%.

La **varianza del error intrasujeto** $(\boldsymbol{\sigma}(\mathbf{e_{tij}}))$ se fija en el primer problema de investigación a: 0%, 0,5%, 5%, 20%, 33,3%, 35%, 50%, 65%, 95% y 100% y en el segundo a un 20%. En esta simulación no se fijan errores de medida en cada una de las nueve ocasiones, sino que el error es aleatorio en cada aplicación, y lo que se fija es la variabilidad del mismo.

Con respecto a la **correlación entre el estatus y la pendiente a nivel de aula** $(\boldsymbol{R_{aulas}})$ y a **nivel de estudiante** $(\boldsymbol{R_{estudiantes}})$, en casi todos los modelos del primer problema de investigación, se han definido cuatro valores. Dos valores positivos 0,2 y 0,9 y dos valores negativos -0,2 y -0,9, excepto en uno de los modelos simulados, ya que la correlación entre el estatus y la pendiente del nivel de aula se fija -0,15. En el segundo problema de investigación también la correlación entre el intercepto y la pendiente del nivel de aula se fija -0,15, y en el nivel de estudiantes a 0,2.

Cada uno de estos valores tiene una doble interpretación. Por un lado, se hace referencia a la fuerza de relación entre el intercepto y la pendiente y por otro, al sentido de la relación entre ambos parámetros (positivo o negativo).

En cuanto a la intensidad de relación se ha optado por valores que representan una relación entre el intercepto y la pendiente baja (0,2) y alta (0,9), ya que siguiendo a Wonnacott y Wonnacott (1999), se considera que un valor de correlación menor a 0,20 es muy bajo, entre 0,20 y 0,40 bajo, entre 0,40 y 0,70 moderado, entre 0,70 y 0,90 alto, y mayor a 0,90 muy alto. Fijar un valor de correlación bajo (0,2) entre el intercepto y la pendiente, supone que el nivel de rendimiento promedio obtenido en la primera aplicación, no se encuentra muy relacionado con el adquirido a lo largo del tiempo. Por el contrario, fijar niveles altos de correlación (0,9), indican que el rendimiento promedio obtenido en la primera ocasión de medida, se encuentra muy relacionado con el que se alcanza a lo largo del resto de aplicaciones.

Referido al sentido de la relación, si el valor de correlación tiene una direccionalidad positiva, indica que cuando el rendimiento de partida es alto, dicho rendimiento también lo será a lo largo del tiempo, mientras que cuando se parten de niveles de rasgo que son bajos, este rendimiento no se incrementará mucho a lo largo del resto de aplicaciones de medida. Sin embargo, si la correlación es negativa, supone que cuando se parte de niveles de rendimiento bajos, este se incrementará mucho a lo largo de las mediciones realizadas, pero si se parte de un rendimiento alto, el mismo se mantiene estable a lo largo del tiempo o crecerá poco.

Finalmente, tal y como se ha comentado, el **rendimiento promedio en la primera aplicación** ($\boldsymbol{\beta_{000}}$), es un valor constante que se ha fijado siempre a 4,6. Del mismo modo, el **rendimiento promedio a lo largo del tiempo** ($\boldsymbol{\beta_{100}}$), también es una constante que se fija en las distintas situaciones simuladas a 0,4.

Los valores de estos coeficientes fijos, al igual que los valores de los coeficientes aleatorios, se encuentran en una escala que van de 0 a 10 y se interpretan en términos de porcentaje. De esta manera el valor de 0 representa el 0% de la varianza y el 10, el 100% de la varianza. Aunque lo importante no son los valores que se fijan en las varianzas, si no la diferencia existente entre los porcentajes de varianza fijados en cada uno de los coeficientes, y el porcentaje fijado en cada coeficiente respecto a la varianza total del modelo.

Con el fin de resumir y clarificar estas condiciones experimentales, a continuación, se presenta una tabla (Tabla V.1) en la cual, se especifica cada una de las características manipuladas con sus valores correspondientes, dentro de los distintos problemas de investigación.

Tabla V.1. *Condiciones experimentales del primer y segundo problema de investigación*

	Primer Problema de Investigación	Segundo problema de Investigación
Condición	**Valor**	
N_{aulas}	400	*Situación a:* 942 (75% aulas pequeñas) *Situación b*: 654 (75% aulas medianas) *Situación c*: 464(75% aulas grandes)
$N_{estudiantes\_total}$	10.000	10.000
$N_{estudiantes\_mínimo}$	20	*Aulas pequeñas:*5 *Aulas medianas:*11 *Aulas grandes:*20
$N_{estudiantes\_máximo}$	30	*Aulas pequeñas:10* *Aulas medianas:19* *Aulas grandes:30*
β_{000}	4,6	4,6
β_{100}	0,4	0,4
$\sigma(u_{00j})$	0%, 1,25%, 2,5%, 3,36%, 4,35%, 5%, 12,5%, 15%, 23, 75%, 33,3%, 40%, 55% y 95%	15%
$\sigma(u_{10j})$	0%, 1,25%, 5%, 10%, 12,5%, 12,27%, 23, 75%, 25%, 33,3%, 47,5% y 95%	5%
$\sigma(r_{0ij})$	0%, 1,25%, 2,5%, 5%, 10%, 12,5%, 15%, 17,5%, 23, 75%, 25%, 33,3%, 35%, 37,7%, 40%, 47,5%, 48,96%50%, 65%, 85%,95% y 100%	35%
$\sigma(r_{1ij})$	0%, 1%, 1,25%, 2,5%, 5%, 6,49%, 8,42%, 10%, 12,5%, 17,5%, 23, 75%, 25%, 33,3%, 35%, 40%, 47,5% ,50%, 65%, 95% y 100%,	25%
$\sigma(e_{tij})$	0%, 0,5%, 5%, 20%, 33,3%, 35%, 50%, 65%, 95% y 100%	20%
R_{aulas}	0,2; 0,9; -0,2; -0,9; -0,15	-0,15
$R_{estudiantes}$	0,2; 0,9; -0,2; -0,9; -0,15	0,2

V.3.2.2. Determinación de las variables dependientes

Las variables dependientes están condicionadas por los objetivos de la investigación. Por ello, en el primer problema de investigación, centrado en el estudio del efecto de la variabilidad y la correlación entre el intercepto y la pendiente, en la fiabilidad con la que se estima los niveles de logro de los estudiantes a lo largo del tiempo, la variable dependiente es **la matriz de correlaciones generada entre las puntuaciones de los estudiantes, en las distintas ocasiones de medida**, tras la manipulación de las diferentes condiciones experimentales.

En el segundo problema de investigación, donde se pretende estudiar si afecta el tamaño de las aulas en la estimación de la pendiente de crecimiento de las mismas, la variable dependiente es el **residuo aleatorio de la pendiente de crecimiento del nivel aulas** ($\mathbf{u_{00j}}$)

V.3.3. Plan de análisis de datos

Como ya se ha comentado en los apartados precedentes, la simulación Monte-Carlo se centra en generar múltiples muestras que se corresponden a unos parámetros previamente establecidos.

Para la realización de las simulaciones de dicho capítulo, se ha utilizado el programa R-Studio versión 3.4.3 (R Core Team, 2017). Mediante dicho software gratuito, se programa un modelo anidado con tres niveles (tiempo, estudiante y aula) y longitudinal, nueve ocasiones de medida, donde el número de réplicas realizadas para cada situación simulada es de 1.000, ya que el error que se comete al aproximar una distribución mediante simulación, es inversamente proporcional al número de replicaciones realizadas (Diaz- Emparanza, 1995), y la idea es que exista el menor error posible.

También se ha utilizado el programa IBM SPSS versión 21, para la computación de algunos resultados, así como el EXCEL, tal y como se explicará mas adelante.

A continuación, se procede a la descripción de los pasos que se han seguido para la generación de los datos.

V.3.3.1. Procedimiento para la generación de datos

Este trabajo de investigación está basado en una simulación, por ello, lo primero que se hizo fue la elaboración de un procedimiento adecuado que permitiera la generación de las distintas condiciones experimentales deseadas. La calidad de los resultados de la simulación depende del número de replicaciones, de la precisión numérica de los programas de generación y del análisis de datos (Stone & Norman, 1993), aspectos que se tuvieron en consideración a lo largo de todo el procedimiento.

Dentro del ***primer problema de investigación***, para dar respuesta a cada una de las tres situaciones planteadas en el segundo objetivo específico (ver apartado *V.2. Objetivos de Investigación*) se desarrollan tres scripts diferentes, que aluden a cada una de las situaciones. A modo de ejemplo, se explica el script referido a la *situación a,* en la cual sólo se fijan valores de varianza en algunos de los coeficientes aleatorios del modelo. El resto de scripts aparecen explicados en el Anexo I del pen drive, en el cual se adjuntan los scripts para ser ejecutados directamente en *R-studio*.

Tal y como se puede observar en el cuadro 1, el primer paso para la generación de la simulación en el *programa R*, es la instalación de los paquetes necesarios que permiten la realización de la misma.

El paquete MASS (Brian et al., 2017) contiene la función *mvrnorm* que sirve para crear muestras con una distribución normal multivariada, el paquete Matrix (Bates & Maechler, 2017) contiene funciones para trabajar con matrices, y finalmente el paquete Corrplot (Wei et al., 2017) permite crear gráficos basados en matrices de correlaciones.

Además, antes de comenzar con la simulación propiamente dicha, es importante fijar el directorio de trabajo con el fin de guardar los resultados en una carpeta determinada y, por otro lado, establecer una semilla para que, como ya se ha comentado, se generen bajo las mismas condiciones, los mismos valores aleatorios.

Cuadro 1: Pasos previos a la simulación

```
#Instalación de librerías o paquetes
install.packages("MASS")
install.packages("Matrix")
install.packages("corrplot")

#Cargar paquetes
library("MASS")
library("Matrix")
library("corrplot")

#Fijar directorio
setwd("C:/Users/MESE_1/Desktop/prueba/prueba a")

#Fijar semilla
set.seed(123)
```

A continuación, se explica como se ha generado la muestra.

Dentro del primer problema de investigación, tal y como se observa en el cuadro 2, el número total de estudiantes fijados son 10.000, que se reparten en aulas comprendidas por un mínimo de 20 estudiante y un máximo de 30, siendo la media 25.

Cuadro 2: Simulación de la muestra

```
#Generación de la muestra
Nsuj <- 10000# número de estudiantes
TotalGrupos <- Nsuj %/% 25 + (Nsuj %/% 50) #Media y el doble de la media

i<-0
Salir <- FALSE
Resto <- Nsuj
MaxTam <- 30
X <- rep(0,TotalGrupos ) #TotalGrupos esta calculado para que sea más grande de lo
#necesario y no de error
```

Para la generación de la muestra se ha decidido utilizar una distribución binomial, donde tal y como se observa en el cuadro 3, requiere de una serie de argumentos:

- Especificación del número mínimo de estudiantes por aula: 20 estudiantes.
- Especificación de cómo se va a incrementar la muestra: en este caso de manera consecutiva, es decir, de uno en uno.
- Tamaño máximo muestral: el tamaño máximo de estudiantes por aula que se desean es 30, por ello se pone 10, ya que el mínimo se ha fijado a 20
- Probabilidad de éxito en cada prueba: 0,5.

Mediante un bucle (while) se va generando un número aleatorio de sujetos en cada grupo, hasta que el número restante es menor o igual a dos veces el tamaño máximo del grupo que se ha decidido, en ese caso, el resto se divide entre dos.

Cuadro 3: Continuación de la simulación de la muestra

```
while (Salir == FALSE) {
 i<-i+1;
 if (Resto <= MaxTam*2){
  X[i] <- Resto %/% 2;
  Resto <- Resto - X[i];
  X[i+1] <- Resto;
  Salir <- TRUE }
 else
 {X[i] <- 20+rbinom(1,10,0.5); #Caracteristicas de la binomial
 Resto <- Resto - X[i]};
 if (Resto<=0) {Salir <- TRUE}
} #Aquí termina el bucle While
```

Finalmente, dentro de la generación de la muestra, se eliminan los elementos que valen 0 y se comprueba que todo el proceso es correcto, sumando el número total de estudiantes por aula, realizando una tabla en la que se especifica el número de estudiantes por clase, seleccionando el número total de aulas, y realizando un histograma de la distribución muestral.

Cuadro 4: Confirmación de que se ha generado la muestra correctamente

```
X <- X[which(X != 0)] #Aquí se eliminan los elementos de X del final que
#valen 0
X
sum(X) #Esto tiene que ser igual a Nsuj
table(X) #Especifica el n de estudiantes por aula
Nvalues <- length(unique(X))
Ngrupos <- length(X) #Número total de aulas
Nvalues
Ngrupos
hist(X, col="yellow", main= "Histograma de distribución muestral",
   xlab="Número de estudiantes por aula", ylab="Frecuencias")
```

El paso siguiente, como se puede apreciar en el cuadro 5, es la especificación de los valores de las variables independientes o condiciones experimentales. Hay que hacer notar al lector, que, aunque inicialmente se denominó al tercer nivel escuelas, hace referencia a las aulas.

Cuadro 5: Generación de variables independientes

```
#Generación de variables
nstudents = X
nschools = Ngrupos
nrow = Nsuj
vst_k = 1.5 #Varianza del estatus del nivel tres
vp_k = 0.5 #Varianza de la pendiente del nivel tres
vst_j <- 3.5 #Varianza del intercepto del nivel dos
vp_j <- 2.5 # Varianza de la pendiente del nivel dos
vt_j<- 2 #Varianza del error intrasujeto
b0 = 4.6 #Media general del estatus
b1 = 0.4 #Media general de la pendiente
cork = (-0.15) # Correlación entre el intercepto y la pendiente del nivel tres
corj = (0.20) # Correlación entre el intercepto y la pendiente del nivel dos
n = length(cork)*length(corj)
```

A continuación, se crea una tabla llamada resultdf donde se guardan todos los datos generados a partir de los valores fijados en las variables independientes.

Cuadro 6: Generación de data. frame resultdf

```
#Generación del data.frame con todos los resultados
resultdf = data.frame('cork' = numeric(n),
                      'cork_est' = numeric(n),
                      'corj' = numeric(n),
                      'corj_est' = numeric(n),
                      'vst_k' = numeric(n),
                      'vst_k_est' = numeric(n),
                      'vp_k' = numeric(n),
                      'vp_k_est' = numeric(n),
                      'vst_j' = numeric(n),
                      'vst_j_est' = numeric(n),
                      'vp_j' = numeric(n),
                      'vp_j_est' = numeric(n),
                      'vt_j' = numeric(n),
                      'vt_j_0ij_est' = numeric(n),
                      'vt_j_1ij_est' = numeric(n),
                      'vt_j_2ij_est' = numeric(n),
                      'vt_j_3ij_est' = numeric(n),
                      'vt_j_4ij_est' = numeric(n),
                      'vt_j_5ij_est' = numeric(n),
                      'vt_j_6ij_est' = numeric(n),
                      'vt_j_7ij_est' = numeric(n),
                      'vt_j_8ij_est' = numeric(n)
                      ) #Aquí se cierra el data.frame
```

Posteriormente se define el número de réplicas, y se crean diferentes listas para guardar los resultados finales: las puntuaciones de los estudiantes en las diferentes ocasiones de medida en cada una de las réplicas (mydata_scores), las correlaciones entre las puntuaciones de los estudiantes en las diferentes ocasiones de medida, en las 1000 réplicas (mydata_corr_occasions) y, los errores de esas matrices de correlación (mydata_corr_errors).

Además, se crea una lista denominada mdresult con el fin de guardar los valores obtenidos en cada una de las variables independientes generadas en las 1000 réplicas.

Cuadro 7: Generación de objetos y especificación de réplicas

```
#Generación de la listas para los resultados finales
mydata_scores<-list()
mydata_corr_occasions<-list()
mydata_corr_errors<-list()
#Especificación del número de réplicas
replicas<-1000
#Generación de una lista para guardar todos los resultados
mdresult<-list()
```

En los siguientes cuadros se muestra la generación de los distintos parámetros del modelo multinivel.

Cuadro 8: Generación de varianza y covarianza del nivel tres y dos

```
#Generación de las varianzas y covarianzas para las 1000 réplicas
for(l in 1:replicas){ #Se abre un bucle para las réplicas
 row_res = 1 #Va desde el valor 1 hasta el máximo de combinaciones
#de correlación dentro de cada réplica

for(k in cork){ #Se abre un bucle para la correlación entre la varianza del
#intercepto y la pendiente del nivel tres
   cov_k = k*sqrt(vst_k)*sqrt(vp_k)
   sigma_u = matrix(data=c(vst_k,cov_k,cov_k,vp_k),nrow=2,ncol=2,byrow=T)

 for(j in corj){ #Se abre un bucle entre la varianza del intercepto y la
pendiente del nivel #dos
    cov_j = j*sqrt(vst_j)*sqrt(vp_j)
    sigma_r = matrix(data=c(vst_j,cov_j,cov_j,vp_j),nrow=2,ncol=2,byrow=T)

    row = 1 #Valores de 1 al número de estudiantes por escuela o aula y se
inicia en cada #réplica de cada combinación
```

Cuadro 9: Generación de los efectos de las aulas o escuelas y estudiantes

```
    mydf = data.frame(school = numeric(nrow), #Generación de efectos
aleatorios
                student = numeric(nrow),
                u00J = numeric(nrow),
                u10J = numeric(nrow),
                r0IJ = numeric(nrow),
                r1IJ = numeric(nrow)

    )
```

```
for(school in 1:nschools){ #Bucle de los efectos de la escuela
      u = mvrnorm(1, c(0,0), sigma_u) #Efectos aletorios de las aula, u = (u00j,
u10j)

   for(student in 1:nstudents[school]){ #Dentro de las escuelas se asigna a los
estudiantes
        r = mvrnorm(1, c(0,0), sigma_r) #Efectos aleatorios de los estudiantes, r
= (r0ij, r01j)
```

```
        mydf$school[row] = school
        mydf$student[row] = student
        mydf$u00J[row] = u[1]
        mydf$u10J[row] = u[2]
        mydf$r0IJ[row] = r[1]
        mydf$r1IJ[row] = r[2]
        row = row + 1 #Valores de 1 al número de estudiantes por escuela y se
inicia en cada #réplica de cada combinación
   } #Cierro el bucle referido a los estudiantes for(student)

  } #Cierro el bucle referido a la escuela for(school)
```

Cuadro 10: Comprobación de los valores de correlación entre el intercepto y la pendiente y de los valores de varianza (se presentan los fijados y estimados)

```
#Valores fijados y estimados
      resultdf$cork[row_res] = k
      resultdf$cork_est[row_res] = cor(mydf$u00J,mydf$u10J, method = "pearson")
#Correlación #nivel tres
      resultdf$corj[row_res] = j
      resultdf$corj_est[row_res] = cor(mydf$r0IJ,mydf$r1IJ,method = "pearson" )#Correlación
#nivel dos
      resultdf$vst_k[row_res] = vst_k #Varianza del intercepto del nivel tres
      resultdf$vst_k_est[row_res] = var(mydf$u00J) #Varianza del intercepto estimado del
#nivel tres
      resultdf$vp_k[row_res] = vp_k #Varianza de la pendiente del nivel tres
      resultdf$vp_k_est[row_res] = var(mydf$u10J) #Varianza de la pendiente del nivel tres
#estimado
      resultdf$vst_j[row_res] = vst_j #Varianza del estatus del nivel dos
      resultdf$vst_j_est[row_res] = var(mydf$r0IJ) #Varianza del estatus del nivel dos
estimado
      resultdf$vp_j[row_res] = vp_j#varianza de la pendiente del nivel dos
      resultdf$vp_j_est[row_res] = var(mydf$r1IJ) #Varianza de la pendiente del nivel dos
#estimado
```

Cuadro 11: Se añaden los residuos y se calculan las puntuaciones de los estudiantes

```
      #Añado el error para cada punto temporal
      errors = as.data.frame(mvrnorm(nrow, rep(0,9), diag(rep(vt_j,9))))
      colnames(errors) = c('e_0ij', 'e_1ij', 'e_2ij', 'e_3ij', 'e_4ij', 'e_5ij','e_6ij', 'e_7ij', 'e_8ij')
      mydf = merge(mydf, errors, by = 'row.names')
      mydf = mydf[,-1]
```

```
#Ordeno a las escuelas y a los estudiantes dentro de las escuelas o aulas
      mydf <- mydf[order(+mydf$school, mydf$student), ]
      write.table(mydf ,file='mydf
.csv',quote=FALSE,col.names=TRUE,row.names=FALSE,sep=';',dec=',')
```

```
# Calculo de las puntuaciones en cada momento temporal
    mydf$Y_0ij = b0 + b1*0 + mydf$r0IJ + mydf$r1IJ*0 + mydf$u00J + mydf$u10J*0 +
mydf$e_0ij
    mydf$Y_1ij = b0 + b1*1 + mydf$r0IJ + mydf$r1IJ*1 + mydf$u00J + mydf$u10J*1 +
mydf$e_1ij
    mydf$Y_2ij = b0 + b1*2 + mydf$r0IJ + mydf$r1IJ*2 + mydf$u00J + mydf$u10J*2 +
mydf$e_2ij
    mydf$Y_3ij = b0 + b1*3 + mydf$r0IJ + mydf$r1IJ*3 + mydf$u00J + mydf$u10J*3 +
mydf$e_3ij
    mydf$Y_4ij = b0 + b1*4 + mydf$r0IJ + mydf$r1IJ*4 + mydf$u00J + mydf$u10J*4 +
mydf$e_4ij
    mydf$Y_5ij = b0 + b1*5 + mydf$r0IJ + mydf$r1IJ*5 + mydf$u00J + mydf$u10J*5 +
mydf$e_5ij
    mydf$Y_6ij = b0 + b1*6 + mydf$r0IJ + mydf$r1IJ*6 + mydf$u00J + mydf$u10J*6 +
mydf$e_6ij
    mydf$Y_7ij = b0 + b1*7 + mydf$r0IJ + mydf$r1IJ*7 + mydf$u00J + mydf$u10J*7 +
mydf$e_7ij
    mydf$Y_8ij = b0 + b1*8 + mydf$r0IJ + mydf$r1IJ*8 + mydf$u00J + mydf$u10J*8 +
mydf$e_8ij
```

```
    #Obtengo los errores de mydf y los pongo en resultdf que es un date frame
    resultdf$vt_j[row_res] = vt_j
    resultdf$vt_j_0ij_est[row_res]= var(mydf$e_0ij)
    resultdf$vt_j_1ij_est[row_res]= var(mydf$e_1ij)
    resultdf$vt_j_2ij_est[row_res]= var(mydf$e_2ij)
    resultdf$vt_j_3ij_est[row_res]= var(mydf$e_3ij)
    resultdf$vt_j_4ij_est[row_res]= var(mydf$e_4ij)
    resultdf$vt_j_5ij_est[row_res]= var(mydf$e_5ij)
    resultdf$vt_j_6ij_est[row_res]= var(mydf$e_6ij)
    resultdf$vt_j_7ij_est[row_res]= var(mydf$e_7ij)
    resultdf$vt_j_8ij_est[row_res]= var(mydf$e_8ij)
```

Las puntuaciones de los estudiantes en cada ocasión de medida son guardadas en una tabla denominada score. A partir de las mismas, se calcula las correlaciones entre las puntuaciones de los estudiantes en las distintas ocasiones de medidas y se guardan en un objeto denominado corr_occasions. Los errores de esas correlaciones se guardan en el objeto corr_errors. Asimismo, en el objeto mdresult se guardan los valores originales que se han fijado en las variables independientes, y los valores que se han obtenido en cada una de las 1000 réplicas.

Si se quiere consultar el valor real y el obtenido en cada una de las réplicas, solo hay que ejecutar la función mdresult [[l]] poniendo dentro de los corchetes el número de réplica que se quiere consultar.

Cuadro12: Se guardan resultados

```
#Guardo los resultados de las puntuaciones de los estudiantes en cada momento
#temporal
      scores = mydf[,c('Y_0ij', 'Y_1ij', 'Y_2ij', 'Y_3ij', 'Y_4ij', 'Y_5ij','Y_6ij', 'Y_7ij', 'Y_8ij')]
#Guardo los resultados de la matriz de correlaciones de las puntuacines de los
#estudiantes
      corr_occasions = cor(mydf[,c('Y_0ij', 'Y_1ij', 'Y_2ij', 'Y_3ij', 'Y_4ij', 'Y_5ij','Y_6ij', 'Y_7ij',
'Y_8ij')])
#Guardo los resultados de la matriz residual
      corr_errors = cor(mydf[,c('e_0ij', 'e_1ij', 'e_2ij', 'e_3ij', 'e_4ij', 'e_5ij','e_6ij', 'e_7ij', 'e_8ij')])
```

```
#Redondeo
      mydata_scores[[(row_res-1)*replicas+l]]<-round(scores,4)
      mydata_corr_occasions[[(row_res-1)*replicas+l]]<-round(corr_occasions,4)
      mydata_corr_errors[[(row_res-1)*replicas+l]]<-round(corr_errors,4)
      row_res = row_res + 1 #Va desde el valor 1 hasta el máximo de combinaciones de
#correlación dentro de cada réplica.
```

```
 } #Se cierra el bucle de la correlación entre la varianza y la pendiente del nivel de
#estudiante (CORK)
 } # Se cierra el bucle entre la varianza del intercepto y la pendiente del nivel de escuela
(CORJ)

assign(paste("scores_rep",l,sep=""),mydata_scores)
assign(paste("corr_rep",l,sep=""),mydata_corr_occasions)
assign(paste("errors_rep",l,sep=""),mydata_corr_errors)
```

```
mdresult[[l]]<-resultdf #Meto los resultados de resultdf en mdresult, que lo va hacer por
#replicas

} #Se cierra el bucle de la réplicas

mdresult #Todos los resultados en todas las réplicas
mdresult[[1]]#Pido los resultados por réplica
```

Como se realizan un gran número de réplicas (1.000), con el fin de sintetizar los resultados de las mismas y conocer si existe una adecuada recuperación de los valores originales fijados, se calcula el valor promedio de cada uno de los coeficientes obtenidos en las 1.000 réplicas. Estos valores se guardan en un objeto denominado mean_resultdf. En este objeto aparece el valor real fijado en cada coeficiente aleatorio y el valor promedio obtenido en las 1000 réplicas.

En los siguientes cuadros se muestra el cálculo de los valores promedios.

Cuadro 13: Cálculo de los valores promedios

```
#A continuación, se sacan las medias de los valores estimados en las distintas réplicas
sum_mdresult2<-numeric(1);sum_mdresult4<-numeric(1);sum_mdresult6<-numeric(1); #El 1
#cambiar de acuerdo al número de combinacionesde correlación
sum_mdresult8<-numeric(1);sum_mdresult10<-numeric(1);sum_mdresult12<-numeric(1);
sum_mdresult14<-numeric(1);sum_mdresult15<-numeric(1);sum_mdresult16<-numeric(1);
sum_mdresult17<-numeric(1);sum_mdresult18<-numeric(1);sum_mdresult19<-numeric(1);
sum_mdresult20<-numeric(1);sum_mdresult21<-numeric(1);sum_mdresult22<-numeric(1)
```

```
for(h in 1:replicas){
 sum_mdresult2<-sum_mdresult2+mdresult[[h]][,2]
 sum_mdresult4<-sum_mdresult4+mdresult[[h]][,4]
 sum_mdresult6<-sum_mdresult6+mdresult[[h]][,6]
 sum_mdresult8<-sum_mdresult8+mdresult[[h]][,8]
 sum_mdresult10<-sum_mdresult10+mdresult[[h]][,10]
 sum_mdresult12<-sum_mdresult12+mdresult[[h]][,12]
 sum_mdresult14<-sum_mdresult14+mdresult[[h]][,14]
 sum_mdresult15<-sum_mdresult15+mdresult[[h]][,15]
 sum_mdresult16<-sum_mdresult16+mdresult[[h]][,16]
 sum_mdresult17<-sum_mdresult17+mdresult[[h]][,17]
 sum_mdresult18<-sum_mdresult18+mdresult[[h]][,18]
 sum_mdresult19<-sum_mdresult19+mdresult[[h]][,19]
 sum_mdresult20<-sum_mdresult20+mdresult[[h]][,20]
 sum_mdresult21<-sum_mdresult21+mdresult[[h]][,21]
 sum_mdresult22<-sum_mdresult22+mdresult[[h]][,22]
}#Aqui termina el bucle h
```

```
#Aquí se hacen las medias
mean_mdresult2<-sum_mdresult2/replicas;mean_mdresult4<-
sum_mdresult4/replicas;mean_mdresult6<-sum_mdresult6/replicas;
mean_mdresult8<-sum_mdresult8/replicas;mean_mdresult10<-
sum_mdresult10/replicas;mean_mdresult12<-sum_mdresult12/replicas;
mean_mdresult14<-sum_mdresult14/replicas;mean_mdresult15<-
sum_mdresult15/replicas;mean_mdresult16<-sum_mdresult16/replicas;
mean_mdresult17<-sum_mdresult17/replicas;mean_mdresult18<-
sum_mdresult18/replicas;mean_mdresult19<-sum_mdresult19/replicas;
mean_mdresult20<-sum_mdresult20/replicas;mean_mdresult21<-
sum_mdresult21/replicas;mean_mdresult22<-sum_mdresult22/replicas
```

```
#Creación del data.frame para guardar los resultados promedios
mean_resultdf<-data.frame('cork'=mdresult[[1]][,1],
                'cork_est' = mean_mdresult2,
                'corj' = mdresult[[1]][,3],
                'corj_est' = mean_mdresult4,
                'vst_k' = mdresult[[1]][,5],
                'vst_k_est' = mean_mdresult6,
                'vp_k' = mdresult[[1]][,7],
                'vp_k_est' = mean_mdresult8,
                'vst_j' = mdresult[[1]][,9],
                'vst_j_est' = mean_mdresult10,
                'vp_j' = mdresult[[1]][,11],
                'vp_j_est' = mean_mdresult12,
                'vt_j' = mdresult[[1]][,13],
                'vt_j_0ij_est' = mean_mdresult14,
                'vt_j_1ij_est' = mean_mdresult15,
                'vt_j_2ij_est' = mean_mdresult16,
                'vt_j_3ij_est' = mean_mdresult17,
                'vt_j_4ij_est' = mean_mdresult18,
                'vt_j_5ij_est' = mean_mdresult19,
                'vt_j_6ij_est' = mean_mdresult20,
                'vt_j_7ij_est' = mean_mdresult21,
                'vt_j_8ij_est' = mean_mdresult22
```

```
#Media del promedio de los errores y lo añado a la tabla de mean_resultdf

mean_errores <-
(mean_resultdf$vt_j_0ij_est+mean_resultdf$vt_j_1ij_est+mean_resultdf$vt_j_2ij_est+
        mean_resultdf$vt_j_3ij_est+ mean_resultdf$vt_j_4ij_est+mean_resultdf$vt_j_5ij_est+
       mean_resultdf$vt_j_6ij_est+mean_resultdf$vt_j_7ij_est+mean_resultdf$vt_j_8ij_est)/9
```

```
mean_resultdf<-data.frame('cork'=mdresult[[1]][,1],
                'cork_est' = mean_mdresult2,
                'corj' = mdresult[[1]][,3],
                'corj_est' = mean_mdresult4,
                'vst_k' = mdresult[[1]][,5],
                'vst_k_est' = mean_mdresult6,
                'vp_k' = mdresult[[1]][,7],
                'vp_k_est' = mean_mdresult8,
                'vst_j' = mdresult[[1]][,9],
                'vst_j_est' = mean_mdresult10,
                'vp_j' = mdresult[[1]][,11],
                'vp_j_est' = mean_mdresult12,
                'vt_j' = mdresult[[1]][,13],
                'vt_j_0ij_est' = mean_mdresult14,
                'vt_j_1ij_est' = mean_mdresult15,
                'vt_j_2ij_est' = mean_mdresult16,
                'vt_j_3ij_est' = mean_mdresult17,
                'vt_j_4ij_est' = mean_mdresult18,
                'vt_j_5ij_est' = mean_mdresult19,
                'vt_j_6ij_est' = mean_mdresult20,
                'vt_j_7ij_est' = mean_mdresult21,
                'vt_j_8ij_est' = mean_mdresult22,
                'vt_j_prom_est' = mean_errores
```

Además, también se calculan la diferencia entre el valor real y el valor promedio obtenido (sesgo)

Cuadro 14: Cálculo del sesgo

```
#Se calcula el sesgo (diferencia entre el valor fijado y el obtenido)
resultdf$cork_diff = mean_resultdf$cork_est - resultdf$cork
resultdf$corj_diff = mean_resultdf$corj_est - resultdf$corj
resultdf$vst_k_diff = mean_resultdf$vst_k_est - resultdf$vst_k
resultdf$vp_k_diff = mean_resultdf$vp_k_est - resultdf$vp_k
resultdf$vst_j_diff = mean_resultdf$vst_j_est - resultdf$vst_j
resultdf$vp_j_diff = mean_resultdf$vp_j_est - resultdf$vp_j
resultdf$vt_j_diff0 = mean_resultdf$vt_j_0ij_est - resultdf$vt_j
resultdf$vt_j_diff1 = mean_resultdf$vt_j_1ij_est - resultdf$vt_j
resultdf$vt_j_diff2 = mean_resultdf$vt_j_2ij_est - resultdf$vt_j
resultdf$vt_j_diff3 = mean_resultdf$vt_j_3ij_est - resultdf$vt_j
resultdf$vt_j_diff4 = mean_resultdf$vt_j_4ij_est - resultdf$vt_j
resultdf$vt_j_diff5 = mean_resultdf$vt_j_5ij_est - resultdf$vt_j
resultdf$vt_j_diff6 = mean_resultdf$vt_j_6ij_est - resultdf$vt_j
resultdf$vt_j_diff7 = mean_resultdf$vt_j_7ij_est - resultdf$vt_j
resultdf$vt_j_diff8 = mean_resultdf$vt_j_8ij_est - resultdf$vt_j
resultdf$vt_j_diff_prom_error = mean_resultdf$vt_j_prom_est- resultdf$vt_j
```

Finalmente, dentro del interés por comprobar que existe una adecuada recuperación de los parámetros aleatorios fijados, se calculan los boxplot de los valores obtenidos en las diferentes réplicas, y a partir de los mismos, la desviación típica de esos valores promedios.

Cuadro 15.a: Cálculo de boxplot y la eficiencia de los errores

```
#Realización de los boxplot de los errores
B<-c()
C<-c()
D<-c()
E<-c()
G<-c()
H<-c()
I<-c()
J<-c()
K<-c()

for (i in 1:replicas)
{B[i]<-mdresult[[i]]$vt_j_0ij_est
C[i]<-mdresult[[i]]$vt_j_1ij_est
D[i]<-mdresult[[i]]$vt_j_2ij_est
E[i]<-mdresult[[i]]$vt_j_3ij_est
G[i]<-mdresult[[i]]$vt_j_4ij_est
H[i]<-mdresult[[i]]$vt_j_5ij_est
I[i]<-mdresult[[i]]$vt_j_6ij_est
J[i]<-mdresult[[i]]$vt_j_7ij_est
K[i]<-mdresult[[i]]$vt_j_8ij_est
} #Cierro bucle i
pdf(file="B,C,D,E,G,H,I,J,K.pdf")
boxplot(B,C,D,E,G,H,I,J,K,col="pink", main="Boxplot varianza errores",
      xlab= "Precisión en la recuperación de parámetros",
names= c("e_0ij", "e_1ij", "e_2ij", "e_3ij", "e_4ij", "e_5ij", "e_6ij", "e_7ij", "e_8ij"))
dev.off()
```

```
#Desviaciones típicas
sdcoefB<-sd(B)
sdcoefC<-sd(C)
sdcoefD<-sd(D)
sdcoefE<-sd(E)
sdcoefG<-sd(G)
sdcoefH<-sd(H)
sdcoefI<-sd(I)
sdcoefJ<-sd(J)
sdcoefK<-sd(K)
#Guardar resultados
write.table(sdcoefB,file='sd_error1.csv',quote=FALSE,col.names=TRUE,row.names=FALSE,sep=';',dec
=',')
write.table(sdcoefC,file='sd_error2.csv',quote=FALSE,col.names=TRUE,row.names=FALSE,sep=';',dec
=',')
write.table(sdcoefD,file='sd_error3.csv',quote=FALSE,col.names=TRUE,row.names=FALSE,sep=';',dec
=',')
write.table(sdcoefE,file='sd_error4.csv',quote=FALSE,col.names=TRUE,row.names=FALSE,sep=';',dec
=',')
write.table(sdcoefG,file='sd_error5.csv',quote=FALSE,col.names=TRUE,row.names=FALSE,sep=';',dec
=',')
write.table(sdcoefH,file='sd_error6.csv',quote=FALSE,col.names=TRUE,row.names=FALSE,sep=';',dec
=',')
write.table(sdcoefI,file='sd_error7.csv',quote=FALSE,col.names=TRUE,row.names=FALSE,sep=';',dec=
',')
write.table(sdcoefJ,file='sd_error8.csv',quote=FALSE,col.names=TRUE,row.names=FALSE,sep=';',dec=
',')
write.table(sdcoefK,file='sd_error9.csv',quote=FALSE,col.names=TRUE,row.names=FALSE,sep=';',dec
=',')
```

Cuadro 15.b: Cálculo del boxplot del valor de correlación fijado entre el intercepto y la pendiente del nivel de estudiante y/o aula y del boxplot de la varianza del intercepto y la pendiente de ese mismo nivel o niveles

```
#Boxplot varianza del intercepto, pendiente y error intrasujeto
L<-c()
M<-c()
N<-c()
Ñ<-c()
O<-c()
P<-c()
for (i in 1:replicas)
{L[i]<-mdresult[[i]]$cork_est
M[i]<-mdresult[[i]]$corj_est
N[i]<-mdresult[[i]]$vst_k_est
Ñ[i]<-mdresult[[i]]$vp_k_est
O[i]<-mdresult[[i]]$vst_j_est
P[i]<-mdresult[[i]]$vp_j_est
} #Cierro bucle i
```

```
#Pdf de cada valor de correlación y las varianzas del intercepto y la pendiente
pdf(file="L.pdf");
pdf(file="M.pdf");
pdf(file="N.pdf");
pdf(file="Ñ.pdf");
pdf(file="O.pdf");
pdf(file="P.pdf");
boxplot(L, col=c("yellow"), main="Boxplot", xlab= "Raulas", names = c("Raulas"))
dev.off()
boxplot(M, col=c("yellow"), main="Boxplot", xlab= "Restudiantes", names = c("Restudiantes"))
dev.off()
boxplot(N, col=c("yellow"), main="Boxplot", xlab= "u_00j", names = c("u_00j"))
dev.off()
boxplot(Ñ, col=c("yellow"), main="Boxplot", xlab= "u_10j", names = c("u_10j"))
dev.off()
boxplot(O, col=c("yellow"), main="Boxplot", xlab= "r_0ij", names = c("r_0ij"))
dev.off()
boxplot(P, col=c("yellow"), main="Boxplot", xlab= "r_1ij", names = c("r_1ij"))
dev.off()
```

```
#Desviaciones típicas
sdcoefL<-sd(L)
sdcoefM<-sd(M)
sdcoefN<-sd(N)
sdcoefÑ<-sd(Ñ)
sdcoefO<-sd(O)
sdcoefP<-sd(P)
#Guardo los resultados de las desviaciones típicas
write.table(sdcoefL,file='sd_Rschools.csv',quote=FALSE,col.names=TRUE,row.names=FALSE,sep=';',d
ec=',')
write.table(sdcoefM,file='sd_Rstudents.csv',quote=FALSE,col.names=TRUE,row.names=FALSE,sep=';',
dec=',')
write.table(sdcoefN,file='sd_st_k.csv',quote=FALSE,col.names=TRUE,row.names=FALSE,sep=';',dec=','
)
write.table(sdcoefÑ,file='sd_p_k.csv',quote=FALSE,col.names=TRUE,row.names=FALSE,sep=';',dec=',')
write.table(sdcoefO,file='sd_st_j.csv',quote=FALSE,col.names=TRUE,row.names=FALSE,sep=';',dec=',')
write.table(sdcoefP,file='sd_p_j.csv',quote=FALSE,col.names=TRUE,row.names=FALSE,sep=';',dec=',')
```

En el siguiente cuadro, en los objetos: scores_rep1, corr_rep1, errors_rep1, se encuentran las puntuaciones de los estudiantes, las correlaciones de las puntuaciones y los errores de dichas correlaciones, respectivamente, de cada réplica. El número fijado después de *rep* hace referencia a la réplica correspondiente, por lo tanto, tan sólo hay que cambiar ese número para especificar la réplica que se desea consultar.

Sin embargo, en los objetos, mydata_scores, mydata_corr_occasions y mydata_corr_errors, se encuentran los resultados de las puntuaciones de los estudiantes, de las correlaciones de esas puntuaciones y de los errores de correlación, respectivamente, referidos a las 1.000 réplicas.

Cuadro 16: Presentación de resultados.

```
#Puntuaciones de los estudiantes por réplicas
scores_rep1#hace referencia a la réplica 1 #Hace referencia a la réplica 1

#Correlaciones por réplicas
corr_rep3 #Hace referencia a la réplica 3

#Errores de correlación por réplica
errors_rep1

#Para perdir los resultados en todas las réplicas
mydata_scores #Puntuaciones
mydata_corr_occasions#Correlaciones
mydata_corr_errors #Errores de correlación
```

Por cuestiones de espacio y con el fin de sintetizar, igual que anteriormente se calcula los valores promedios de las puntuaciones de los estudiantes en las nueve ocasiones de medida, las correlaciones promedias de esas puntuaciones y los errores promedios de las correlaciones de las puntuaciones de los estudiantes. Asimismo, también se calculan, las desviaciones promedias de las puntuaciones y de las correlaciones entre las puntuaciones en las 1.000 réplicas.

Cuadro 17.a: Promedio de las puntuaciones de los estudiantes

```
#Media de las puntuaciones de los estudiantes y de las desviaciones típicas
stdmean_score<-list()
mean_scores<-list()

for(nmediascorre in 1:1){
 scores_occ<-matrix(0,10000,9)
 scores_occ2<-matrix(0,10000,9)
 for(l in 1:replicas){
  scores_occ<-scores_occ+get(paste("scores_rep",l,sep=""))[[nmediascorre]]
  scores_occ2<-scores_occ2+(get(paste("scores_rep",l,sep=""))[[nmediascorre]])^2
 } #Cierre del bucle
 std_corr<-(scores_occ2/replicas-scores_occ^2/replicas^2)^0.5
 scores_occ<-scores_occ/replicas
 mean_scores[[nmediascorre]]<-scores_occ
 stdmean_score[[nmediascorre]]<- std_corr

} #Cierra del bucle
mean_scores
stdmean_score
#Guardo los resultados de las medias y de las desviaciones típicas
write.table( mean_scores,file='mean_scores .csv',quote=FALSE,
          col.names=TRUE,row.names=FALSE,sep=';',dec=',')
write.table(stdmean_score ,file='stdmean_score.csv',quote=FALSE,
         col.names=TRUE,row.names=FALSE,sep=';',dec=',')
```

Cuadro 17. b: Promedios de las correlaciones entre las puntuaciones de los estudiantes

```
# Medias de las correlaciones y de las desviaciones típicas
stdmean_corr<-list()
mean_corr<-list()

for(nmediascorr in 1:1){
 corr_occ<-matrix(0,9,9)
 corr_occ2<-matrix(0,9,9)
 for(l in 1:replicas){
  corr_occ<-corr_occ+get(paste("corr_rep",l,sep=""))[[nmediascorr]]
  corr_occ2<-corr_occ2+(get(paste("corr_rep",l,sep=""))[[nmediascorr]])^2

 } #Cierra el bucle
 std_corr<-(corr_occ2/replicas-corr_occ^2/replicas^2)^0.5
 corr_occ<-corr_occ/replicas
 mean_corr[[nmediascorr]]<-corr_occ
 stdmean_corr[[nmediascorr]]<-std_corr
} #Cierra el bucle
mean_corr #Ver los valores medios de la matriz de correlaciones
stdmean_corr#Desviaciones típicas de las correlaciones
#Guardo los resultados
write.table( mean_corr,file='mean_corr .csv',quote=FALSE,
          col.names=TRUE,row.names=FALSE,sep=';',dec=',')
write.table(stdmean_corr ,file='stdmean_corr.csv',quote=FALSE,
         col.names=TRUE,row.names=FALSE,sep=';',dec=',')

#Gráfico de correlaciones
w.cor=as.data.frame.list(mean_corr)
w.cor=as.matrix(w.cor)
```

Cuadro 17.c: Promedios de los errores de correlación

```
#Media de los errores
mean_errors<-list()

for(nmediaserror in 1:1){
 errors_occ<-matrix(0,9,9)
 for(l in 1:replicas){
  errors_occ<-errors_occ+get(paste("errors_rep",l,sep=""))[[nmediaserror]]

 } #Se cierra el bucle
 errors_occ<-errors_occ/replicas
 mean_errors[[nmediaserror]]<-errors_occ

} #Se cierra el bucle
mean_errors
#Guardo resultados
write.table( mean_errors,file='mean_errors
.csv',quote=FALSE,col.names=TRUE,row.names=FALSE,sep=';',dec=',')
```

Para dar respuesta al ***segundo problema de investigación*** se utiliza el script anterior, referido a la *situación a*, del primer problema de investigación. Tan sólo se modifica el procedimiento desarrollado para la generación de la muestra.

A continuación, se presenta como se ha generado la muestra para la *situación a*, donde, recuérdese que se pretende simular un modelo en el que alrededor de un 75% de aulas sean pequeñas, (compuestas por un mínimo de 5 estudiantes y un máximo de 10), el 12,5% sean medianas (el número mínimo de estudiantes es 11 y el máximo 19) y el otro 12,5% aulas grandes con más de 19 estudiantes por aula, siendo el máximo de 30 estudiantes. El script referido a esta situación como los de la *situación b y c* pueden ser consultados en el Anexo I.

Tal y como aparece en el siguiente cuadro, lo primero que se hace, nuevamente, es fijar una semilla para que siempre se generen los mismos números aleatorios. Posteriormente se especifica la proporción de aulas (pequeñas, medianas y grandes) en cada una de las tres situaciones (SituaciónA, SituaciónB y SituaciónC), el número de aulas de cada tipo, dentro de cada situación (TamañosA, TamañosB y Tamaños C) y el número de estudiantes en cada tipo de aulas, referidas igualmente a cada una de las tres situaciones (SujetosA, SujetosB y SujetosC).

Para realizar los cálculos que permiten obtener estos valores, es necesario tener claro cual es el tamaño medio de cada tipo de aula (pequeño, mediano y grande) y el porcentaje que se quiere de cada una de ellas (75%, 12,5% y 12,5%, respectivamente), en el caso de la *situación a*. Así como el número total de

estudiantes (10.000). A partir de ahí, tal y como se explica en el Excel que se adjunta en el Anexo I, es posible realizar todos los cálculos necesarios para simular los tamaños muestrales de cada una de las tres situaciones (a, b y c).

Cuadro 1: Generación de los valores de la muestra

```
set.seed(125)

SituaciónA <- c(0.5287,0.1767,0.2946)#Proporciones de cada aula
SituaciónB <- c(0.0614,0.7339,0.2047)
SituaciónC <- c(0.0435,0.087,0.8696)
Situaciones <- rbind(SituaciónA,SituaciónB,SituaciónC)

TamañosA <- c(706, 118,118)#número de aulas de cada tipo
TamañosB <- c(82, 490,82)
TamañosC <- c(58, 58,348)
Tamaños <-  rbind(TamañosA,TamañosB,TamañosC)

SujetosA <- c(5295, 1770,2950)#Número de estudiantes en cada aula
SujetosB <- c(615, 7350,2050)
SujetosC <- c(435, 870,8700)
Sujetos <- rbind(SujetosA ,SujetosB ,SujetosC)
ListaFinal <- vector("list", 3)
```

En el siguiente cuadro, se muestra el bucle que se ha desarrollado para generar las distintas situaciones donde se reparte a los estudiantes en aulas con tamaños pequeños (P), medianos (M) y grandes (G).

Cuadro 2: Generación de las situaciones a, b y c

```
for (j in 1:3) { # El índice j toma valores de 1 a 3 para señalar cada una de las tres
situaciones
 #Browser()
        y <- sample( c("P","M","G"), 10000, replace=TRUE, prob=Situaciones[j,] )
        prop.table(table(y))
        y
        GrupoP <- which(y == "P")
        GrupoM <- which(y == "M")
        GrupoG <- which(y == "G")
        Grupos  = list(GrupoP, GrupoM, GrupoG)
        w<-rep(0,10000 )
# Aquí he determinado a qué tipo de escuela (por tamaño) pertenece cada uno de los
10000 sujetos
# Antes he determinado (en el Excel) cuántas escuelas hay de cada tipo.
# Ahora el problema es repartir los sujetos de cada tipo (P, M, G) en cada una de las
escuelas.
# Voy extrayendo muestras, una por escuela, y los elimino del grupo grande
LímiteInferior <- 1
```

A continuación, aparece como se ha generado los tres tipos de aula (pequeña, mediana y grande):

Cuadro 3: Generación de los tipos de aula

```
for (i in 1:3) { #Dentro de cada situación, el índice i apunta a cada tipo de escuela, P, M o G

#Primero genero la variable w con 10000 huecos, uno para cada sujeto. En esta variable
#voy a escribir el número de la escuela a la que pertenece cada sujeto.

if (i>1) {LímiteInferior <- LímiteSuperior + 1}
        LímiteSuperior <-  LímiteInferior + Tamaños[[j,i]] -1
```

Posteriormente, se describe el bucle utilizado para generar el tamaño de cada una de las aulas. Se utiliza una distribución uniforme (runif), donde se especifica el número mínimo de estudiantes en cada tipo de aulas y el máximo. Asimismo, también se crea una lista final donde se guardan los tamaños muestrales correspondientes a la *situación a, b y c.*

Cuadro 4: Generación de los tamaños de cada aula y de la lista final correspondiente a los tamaños muestrales de las tres situaciones.

```
for (k in LímiteInferior:LímiteSuperior ) #Con el índice k apunto a cada una de las escuelas
del #tamaño que estoy creando
        {
                if (i==1) {z <- 4+round(runif(1,1,6))} #Pequeñas
                if (i==2) {z <- 10+round(runif(1,1,9))} #Medianas
                if (i==3) {z <- 19+round(runif(1,1,10))} #Grandes
                if (length(Grupos[[i]])>= z)
                {

                        Z <- sample( Grupos[[i]] , z , replace=FALSE)
                        Grupos[[i]] <- Grupos[[i]]  [! Grupos[[i]] %in% Z]
                        w[Z] <- k
                }
                else
                {z <- length(Grupos[[i]])
                 Z <- sample( Grupos[[i]] , z , replace=FALSE)
                 w[Z] <- k
                        Grupos[[i]] <- Grupos[[i]]  [! Grupos[[i]] %in% Z]
                } #Fin del bucle if

                if (k==LímiteSuperior & length(Grupos[[i]])> 0)
                                {  for (l in  Grupos[[i]]){
                 w[l] <- sample( LímiteInferior:LímiteSuperior, 1, replace=FALSE)
                                        Grupos[[i]] <- Grupos[[i]]  [! Grupos[[i]] %in% w[l]]
                                                                } #Fin del bucle l
                                } #Fin del bucle if & length
        } #Fin del bucle k
} #Fin de for i
ListaFinal[[j]] <- w
} #Fin de for j
```

Finalmente, se comprueba que todo el procedimiento funciona correctamente, obteniendo el número total de estudiantes, el número de estudiantes por aula, el número total de aulas y un histograma de la distribución muestral.

Cuadro 5: comprobación de una adecuada generación de la muestra

```
# Esto es para comprobar que sale bien
sum((table(ListaFinal[[1]])))#Número total de estudiantes
table(table(ListaFinal[[1]] ))#Número de estudiantes en cada aula
Ngrupos <- length(table(ListaFinal[[1]]))#Número total de grupos
Ngrupos
hist(table(ListaFinal[[1]]), col="aquamarine", main= "Histograma de la distribución muestral",
xlab="Número de estudiantes por aula",ylab="Frecuencias")
```

V.3.3.2 Precisión en la recuperación de parámetros (Objetivo específico 1 del primer y segundo problema de investigación)

Una vez elaborado el procedimiento para la generación de los datos, antes de la obtención definitiva de los resultados, con el objetivo de controlar algún posible factor distorsionador, se comprueba que exista un adecuado funcionamiento de la programación realizada., ya que determina la validez interna del experimento.

En primer lugar, para garantizar que los tamaños muestrales obtenidos cumplían con las condiciones fijadas, se realizaron diversas comprobaciones

- Cálculo del sumatorio del número total de estudiantes por aula.
- Realización de una tabla en la que se especifica el número de estudiantes por aula.
- Cálculo del número total de aulas.
- Realización de un histograma de la distribución muestral.

Para avalar la precisión de la recuperación de los parámetros aleatorios fijados y de los valores de correlación, se tienen en cuenta en cada uno de los modelos simulados los criterios que aparecen en la siguiente tabla.

Tabla V.2. *Criterios para la comprobación de una adecuada recuperación de los coeficientes aleatorios*

CRITERIOS	INTERPRETACIÓN
Calculo del valor promedio en las 1.000 réplicas de: • La varianza del intercepto del nivel de aula ($\boldsymbol{\sigma}(\mathbf{u_{00j}})$) • La varianza de la pendiente del nivel del aula ($\boldsymbol{\sigma}(\mathbf{u_{10j}})$) • La varianza del estatus del nivel de estudiantes ($\boldsymbol{\sigma}(\mathbf{r_{0ij}})$) • La varianza de la pendiente del nivel de estudiantes $\left(\boldsymbol{\sigma}(\mathbf{r_{1ij}})\right)$ • Las nueve varianzas del error intrasujeto ($\boldsymbol{\sigma}(\mathbf{e_{tij}})$) • La correlación entre el intercepto y la pendiente del nivel de estudiante (***Restudiantes***) • La correlación entre el intercepto y la pendiente del nivel de aula (***Raulas***)	Para que exista una adecuada recuperación, es necesario que dichos valores promedios sean similares a los fijados previamente.
Error típico de la media (DT_media)	Cuanto mayor sea la varianza del error del efecto medio estimado, menos preciso, menos eficiente, es el estimador (Castro,1997).
Sesgo: diferencia entre el valor promedio del estadístico obtenido y el valor fijado de dicho estadístico	Lo ideal es que sea próximo a 0 (nulo), ya que indica similitud entre el estadístico obtenido y el fijado. Si tiene un valor negativo supone una infra estimación del valor paramétrico de referencia. Por el contrario, si tiene un valor positivo se sobreestima el valor del parámetro de referencia.
Coeficiente de variación (CV) en porcentaje, entre el tamaño de la media y la variabilidad de la variable independiente.	Cuanto mayor sea dicho coeficiente mayor heterogeneidad habrá en los valores de dicha variable.
Boxplot de cada una de las condiciones simuladas	Permite analizar como es la dispersión de los puntos con la mediana, los percentiles 25 y 75 y los valores máximos y mínimos.

Los resultados obtenidos en la recuperación de los parámetros muestrales generados en el primer problema de investigación, aparecen en el apartado VI.1.1, referido a *Resultados del objetivo específico uno*. En ese mismo apartado también aparecen la precisión obtenida en la recuperación de los parámetros fijados en los distintos modelos simulados. Como se desarrollan más de 99 modelos, a modo de ejemplo, dentro de cada situación (*situación a, b y c*), se muestran los resultados obtenidos en dos de los modelos que pueden ser de mayor interés, pero la recuperación de parámetros del resto de modelos puede ser consultados en el Anexo II.

En cuanto a la precisión en la recuperación de los parámetros del segundo problema de investigación, en el capítulo VI de resultados, se presenta únicamente la precisión con la que se recuperan las condiciones muestrales fijadas para cada una de las situaciones (*situación a, b y c*) definidas anteriormente, pues, el resto de condiciones manipuladas son las mismas que las definidas en uno de los modelos del primer problema de investigación. Aun así, si el lector lo desea puede consultar la recuperación de estos parámetros en los diferentes modelos en el Anexo II.

V.3.3.3. Procedimiento de generación de modelos (Objetivo específico 2 del primer y segundo problema de investigación.)

Para dar respuesta a cada uno de los problemas de investigación explicados en el *apartado V.1.*, se simulan una serie de modelos bajo diferentes condiciones, que se explican a continuación.

V.3.3.3.1. Generación de modelos del primer problema de investigación

En los modelos desarrollados dentro del primer problema de investigación, tal y como se ha comentado en el apartado de variables, (ver *apartado V.3.2*) el tamaño muestral fijado es siempre el mismo, esto es, 10.000 alumnos repartidos en 400 aulas, que se encuentran comprendidas entre un mínimo de 20 estudiantes y un máximo de 30, aunque la mayoría de las aulas están formadas por 25 estudiantes.

Además, en los diferentes modelos, los niveles de logro de los estudiantes en las distintas ocasiones de medida y, por ende, los valores de correlación obtenidos entre las distintas mediciones realizadas, se encuentran afectados por el nivel o niveles en los que se fija varianza. Si sólo se fija variabilidad en el nivel de estudiantes, el rendimiento de los mismos se debe sobre todo a sus propias características. Según la OECD (2016b) entre las características de los estudiantes que más afectan a su propio rendimiento, se encuentran factores contextuales como su nivel socio-económico, su origen, su estructura familiar, etc; pero también factores vinculados con su educación previa (asistencia o no a preescolar), repetición de cursos anteriores, actitud hacia la escuela y hacia el aprendizaje, sentimiento de pertenencia al grupo, etc.

Si se fija variabilidad sólo en el nivel de aula, el rendimiento de los estudiantes se encontrará muy afectado por las características de las mismas. En este sentido se

ha demostrado que en el logro de los estudiantes juega un papel muy importante: el clima de clase, el número de estudiantes por aula (ratio), la calidad de la infraestructura física, la disponibilidad de material de trabajo, etc. (OECD, 2016b).

Cuando se fija variabilidad en ambos niveles (estudiante y aula) el rendimiento de los estudiantes se encuentra afectado tanto por las características de los alumnos, como por las características de las aulas.

Teniendo en cuentas estas consideraciones, en primer lugar se desarrollan una serie de modelos iniciales en los cuales tan sólo se fija variabilidad en algunos de los coeficientes aleatorios (objetivo específico 2, **situación a**), posteriormente se desarrollan modelos en los cuales, además de fijar una varianza distinta de cero en el intercepto y en la pendiente de un mismo nivel (nivel de aula o nivel de estudiante), se fijan distintos valores de correlación entre el intercepto y la pendiente de dicho nivel (objetivo específico 2, **situación b**). Finalmente se estudian modelos más afines a la realidad educativa, donde además de fijar variabilidad en los coeficientes aleatorios de los tres niveles (tiempo, estudiante y aula), se establecen diferentes valores de correlación entre el intercepto y la pendiente tanto del nivel dos (estudiantes), como del nivel de aulas (nivel tres) (objetivo específico 2, **situación c**).

Para el desarrollo de estos modelos, el rendimiento promedio en la primera aplicación (β_{000}) se decidió fijar a 4,6 y la pendiente de crecimiento promedia (β_{100}) a 0,4. Estos valores se mantienen constantes a lo largo de todas las simulaciones puesto que los mismos no influyen en los valores obtenidos en la matriz de correlaciones, referida a las puntuaciones de los estudiantes en las diferentes ocasiones de medida, que es el objetivo de análisis en este primer problema de investigación.

Los modelos que se presentan a continuación, se organizan en subsituaciones caracterizadas porque en todas ellas se fija variabilidad en los mismos coeficientes, aunque se diferencian en el porcentaje de varianza que se fija en cada uno de ellos.

- ***Modelos generados en la situación a***

Dentro del **objetivo específico 2, *situación a***, se desarrollan los modelos de las siguientes subsituaciones:

Subsituación 1: bajo este contexto se desarrolla un único modelo, **modelo 1**, en el cual no se fija varianza ni en el intercepto, ni en la pendiente del nivel de aula, ni en los coeficientes aleatorios del nivel de estudiantes; por lo tanto, tal y como se muestra en la Tabla V.3, el 100% de la varianza del modelo se fija en el error intrasujeto.

Tabla V.3. *Modelo 1*

Coeficientes	Valores de los coeficientes
Estatus o intercepto (β_{000})	**4,6**
Pendiente (β_{100})	**0,4**
Correlación entre el intercepto y la pendiente en el nivel de aulas (Raulas)	0
Correlación entre el intercepto y la pendiente en el nivel de estudiantes (Resudiantes)	0
Varianza del intercepto en el nivel de aula (u_{00j})	0
Varianza de la pendiente en el nivel de aula (u_{10j})	0
Varianza del intercepto en el nivel de estudiantes (r_{0ij})	0
Varianza de la pendiente en el nivel de estudiantes (r_{1ij})	0
Varianza del error intrasujeto en el nivel de tiempo (e_{tij})	**100%**

En el **modelo 1**, al fijar a cero las varianzas del intercepto de ambos niveles, se asume que no hay diferencias en el rendimiento inicial de los estudiantes (en general, todos los estudiantes parten del mismo nivel de rendimiento previo) y, por lo tanto, en términos promedios todos los estudiantes de una misma aula, así como las aulas entre sí, cuentan con el mismo nivel de logro en la primera ocasión de medida.

Lo mismo ocurre con las pendientes de crecimiento, como no se fija variabilidad ni en la pendiente del nivel dos, ni en la del nivel tres, en general el rendimiento académico de todos los estudiantes crece al mismo ritmo y, todas las aulas cuentan con el mismo nivel de rendimiento promedio a lo largo del tiempo.

Subsituación 2: en este contexto, se presentan una serie de modelos donde se fijan distintos porcentajes en la varianza de la pendiente del nivel de estudiantes, y en la varianza del error intrasujeto, el resto de parámetros se fijan a 0 (no hay variabilidad).

En el **modelo 2**, el porcentaje de varianza de la pendiente del nivel de estudiante, se fija a un 5% de la varianza total del modelo, por lo tanto, la varianza del error intrasujeto se fija a un 95%.

En el **modelo 2.1**, se incrementa el porcentaje de varianza de la pendiente del nivel de estudiantes respecto al modelo anterior, ahora dicha varianza se fija a un 35% de la varianza total del modelo. El incremento de dicha varianza de la pendiente, respecto al modelo anterior (modelo 2), conlleva una disminución del porcentaje de varianza fijado en el error intrasujeto, que en este modelo se fija a un 65% de la varianza total.

Finalmente, en el **modelo 2.2,** la varianza total del mismo, se encuentra repartida de manera equitativa entre la varianza del error intrasujeto (50%) y la varianza de la pendiente del nivel de estudiantes (50%).

Tabla V.4. *Modelo 2, modelo 2.1 y modelo 2.2.*

Coeficientes	Valores de los coeficientes
Estatus o intercepto (β_{000})	**4,6**
Pendiente (β_{100})	**0,4**
Correlación entre el intercepto y la pendiente en el nivel de aulas (Raulas)	0
Correlación entre el intercepto y la pendiente en el nivel de estudiantes (Resudiantes)	0
Varianza del intercepto en el nivel de aula (u_{00j})	0
Varianza de la pendiente en el nivel de aula (u_{10j})	0
Varianza del intercepto en el nivel de estudiantes (r_{0ij})	0
Varianza de la pendiente en el nivel de estudiantes (r_{1ij})	5%/35%/50%
Varianza del error intrasujeto en el nivel de tiempo (e_{tij})	95%/65%/50%

Los modelos de la subsituación 2, representan una realidad educativa en la cual las características de las aulas, no afectan a los niveles de logro previo de los estudiantes, ni a los que adquieren a lo largo del tiempo, ya que no existe variabilidad ni en el estatus ni en la pendiente.

En cuanto a los estudiantes, en general, todos parten de los mismos niveles de logro (la varianza del estatus del nivel de estudiante se fija a 0), pero su rendimiento a lo largo del tiempo es distinto, es decir, no todos los estudiantes crecen por igual a lo largo de las distintas aplicaciones (se fija varianza en la pendiente del nivel de estudiante).

Al fijar solo varianza en la pendiente del nivel de estudiante, los cambios que se produzcan en los niveles de logro a lo largo de las 9 aplicaciones, y, por ende, en las matrices de correlación en las puntuaciones de los estudiantes entre las distintas ocasiones de medida, son consecuencia casi exclusiva de las características sistemáticas de los estudiantes y no del efecto de las aulas, ya que, tal y como se ha comentado, no existe variabilidad en la pendiente de las mismas.

Subsituación 3: en los modelos anteriores de la subsituación 2, el porcentaje fijado en la varianza del error intrasujeto, es superior o igual al porcentaje fijado en la varianza de la pendiente del nivel de estudiante. A continuación, se plantean tres modelos donde el porcentaje fijado en la varianza del error intrasujeto, es inferior al porcentaje fijado en la varianza de la pendiente del nivel de estudiante.

El **modelo 3** que se desarrolla es similar al modelo 2, pero con los porcentajes de varianza fijados en el error intrasujeto y en la pendiente de los estudiantes, invertidos. De esta manera, en dicho modelo, el porcentaje de varianza de la pendiente del nivel de aula se fija a un 95% de la varianza total del modelo, por lo tanto, la varianza del error intrasujeto se fija a un 5%.

En el **modelo 3.1,** el porcentaje de varianza del nivel del tiempo se fija a un 35% de la varianza total del modelo y, la varianza de la pendiente de los estudiantes a un 65% de esa varianza total. Dicho modelo es similar al modelo 2.1, pero con los porcentajes de la varianza del error intrasujeto y de la varianza de la pendiente del nivel de estudiante intercambiados.

Finalmente, dentro de esta subsituación se desarrolla un modelo donde la varianza del error intrasujeto se fija a 0, por lo tanto, toda la variabilidad se fija en la pendiente de los estudiantes (**modelo 3.2**).

Tabla V.5. *Modelo 3, modelo 3.1 y modelo 3.2*

Coeficientes	Valores de los coeficientes
Estatus o intercepto (β_{000})	**4,6**
Pendiente (β_{100})	**0,4**
Correlación entre el intercepto y la pendiente en el nivel de aulas (Raulas)	0
Correlación entre el intercepto y la pendiente en el nivel de estudiantes (Resudiantes)	0
Varianza del intercepto en el nivel de aula (u_{00j})	0
Varianza de la pendiente en el nivel de aula (u_{10j})	0
Varianza del intercepto en el nivel de estudiantes (r_{0ij})	0
Varianza de la pendiente en el nivel de estudiantes (r_{1ij})	**95%/65%/100%**
Varianza del error intrasujeto en el nivel de tiempo (e_{tij})	**5%/35%/0%**

En estos modelos de la subsituación 3, al igual que en los de la subsituación 2, sólo se fija variabilidad en la pendiente del nivel de estudiante y en el error intrasujeto, por lo tanto, sigue sin haber efecto de las características de las aulas sobre los niveles de logro de los estudiantes.

Todos los estudiantes de una misma aula, en general, parten del mismo nivel de rendimiento y, dicho nivel de partida, es el mismo en todas las aulas. No se fija variabilidad en el intercepto de las aulas ni de los estudiantes.

Sin embargo, como si hay variabilidad en las pendientes de crecimiento de los estudiantes, no todos los estudiantes de una misma aula muestran los mismos niveles de logro en las diferentes mediciones. Esto es debido sobre todo a las características de los estudiantes, pero no al efecto de las aulas, ya que no se fija variabilidad en la pendiente del nivel de aula.

Tras desarrollar una serie de modelos donde se fijan distintos porcentajes en la varianza de la pendiente del nivel de estudiantes y en el error intrasujeto, a continuación, se plantean una serie de modelos (subsituación 4 y 5), donde además

de fijar variabilidad en el error intrasujeto, también se fija varianza en el intercepto del nivel dos (estudiantes). El resto de parámetros aleatorios no varían, ya que se fijan a 0.

Subsituación 4: los modelos de este contexto son muy similares a los desarrollados en la subsituación 2. La única diferencia, es que en los modelos de la subsituación 4 se fija variabilidad en el estatus del nivel de estudiante en vez de en la pendiente. Pero los porcentajes fijados en la varianza del error intrasujeto, como la varianza total del nivel de estudiante, es la misma en ambos escenarios.

El **modelo 4** es un modelo donde el porcentaje de varianza del intercepto del nivel de estudiante se fija a un 5% de la varianza total del modelo, y la varianza del error intrasujeto a un 95%. Dichos porcentajes de varianza son los mismos que los fijados en el modelo 2, con la diferencia de que en el modelo 2, el 5% se fija en la varianza de la pendiente en vez del estatus.

En el **modelo 4.1** el porcentaje de varianza fijado en el error intrasujeto, sigue siendo mayor que el fijado en la varianza del estatus del nivel de estudiante. Pero no existe tanta diferencia entre ambos porcentajes como en el modelo 4. En el modelo 4.1, la varianza del error intrasujeto se fija a un 65% de la varianza total del modelo, el resto de varianza (35%), se fija en el intercepto. En este modelo, la varianza total de cada uno de los niveles (aula, estudiante y tiempo) es la misma que la del modelo 2.1, solo difieren, en que en el modelo 2.1 se fija varianza en la pendiente y no en el intercepto.

Finalmente, dentro de esta subsituación se desarrolla el **modelo 4.2** que al igual que en el modelo 2.3, la mitad de la varianza total del modelo se fija en el nivel de estudiante (50%), y la otra mitad en el error intrasujeto (50%). Pero mientras que en el modelo 4.2 el porcentaje de varianza del nivel de estudiante se fija en el estatus, en el modelo 2.3 se fijaba en la pendiente.

Tabla V.6. *Modelo 4, modelo 4.1 y modelo 4.2*

Coeficientes	Valores de los coeficientes
Estatus o intercepto (β_{000})	**4,6**
Pendiente (β_{100})	**0,4**
Correlación entre el intercepto y la pendiente en el nivel de aulas (Raulas)	0
Correlación entre el intercepto y la pendiente en el nivel de estudiantes (Resudiantes)	0
Varianza del intercepto en el nivel de aula (u_{00j})	0
Varianza de la pendiente en el nivel de aula (u_{10j})	0
Varianza del intercepto en el nivel de estudiantes (r_{0ij})	**5%/35%/50%**
Varianza de la pendiente en el nivel de estudiantes (r_{1ij})	0
Varianza del error intrasujeto en el nivel de tiempo (e_{tij})	**95%/65%/50%**

En los modelos de la subsituación 4, no hay efecto ni de las características de los estudiantes ni de las aulas sobre las pendientes de crecimiento. Los estudiantes, en general, parten de distintos niveles de rendimiento previo (se fija variabilidad en el estatus del nivel de estudiante), pero su ritmo de crecimiento a lo largo del tiempo es el mismo (no se fija variabilidad en las pendientes del nivel de estudiantes). Esto supone que los estudiantes de una misma aula, presentan distintos niveles de logro en la primera ocasión de medida, pero sus niveles de logro a lo largo del tiempo son constantes, apenas se diferencian.

En los modelos de la **subsituación 5**, tal y como se planteó en la subsituación 3, pero en este caso referido a la varianza del intercepto del nivel de estudiantes, en vez de a la varianza de la pendiente, se desarrollan una serie de modelos donde existe más porcentaje de varianza en el intercepto del nivel dos que en la varianza del error intrasujeto, lo contrario a lo que ocurría en los modelos de la subsituación 4.

El **modelo 5** es un modelo donde los porcentajes fijados en los distintos niveles son los mismos que en el modelo 3, pero en el nivel de estudiante se fija una varianza distinta de cero en el estatus en vez de en la pendiente. En este modelo 5, la mayor parte de la varianza total del modelo se encuentra fijada en el intercepto del nivel de estudiante, ya que dicho valor se fija a un 95% de la varianza total del

modelo, mientras que la varianza del error intrasujeto es tan solo un 5% de la varianza total del modelo. Estos porcentajes son los mismos que se fijaron en el modelo 4, pero ahora invertidos.

En el **modelo 5.1**, el porcentaje de varianza del nivel del error intrasujeto se fija a un 35% de la varianza total del modelo y, la varianza del intercepto de los estudiantes a un 65%. Por lo tanto, dicho modelo es similar al modelo 4.1, pero con los porcentajes de la varianza del error intrasujeto y de la varianza del intercepto del nivel de estudiante intercambiados. Además, dichos porcentajes fijados en cada uno de los tres niveles de este modelo, son coincidentes con los del modelo 3.1

Finalmente, dentro de esta subsituación se plantea el **modelo 5.2**, donde la varianza del error intrasujeto se fija a 0, por lo tanto, toda la variabilidad del modelo se fija en el intercepto de los estudiantes (100%). Este modelo donde toda la variabilidad se fija en el nivel de estudiante es similar al modelo 3.2, lo único que en este último modelo toda la varianza se fijaba en la pendiente y en el modelo 5.2 se fija en el estatus.

Tabla V.7. *Modelo 5, modelo 5.1 y modelo 5.2*

Coeficientes	Valores de los coeficientes
Estatus o intercepto (β_{000})	**4,6**
Pendiente (β_{100})	**0,4**
Correlación entre el intercepto y la pendiente en el nivel de aulas (Raulas)	0
Correlación entre el intercepto y la pendiente en el nivel de estudiantes (Resudiantes)	0
Varianza del intercepto en el nivel de aulas (u_{00j})	0
Varianza de la pendiente en el nivel de aulas (u_{10j})	0
Varianza del intercepto en el nivel de estudiantes (r_{0ij})	**95%/65%/100%**
Varianza de la pendiente en el nivel de estudiantes (r_{1ij})	0
Varianza del error intrasujeto en el nivel de tiempo (e_{tij})	**5%/35%/0%**

En estos tres modelos de la subsituación 5, las características diferenciales de los estudiantes hacen que, en general, los mismos partan de niveles de rendimiento distintos, sin embargo, con el paso del tiempo su ritmo de crecimiento es constante (no hay variabilidad en las pendientes de crecimiento). Así pues, las mismas características sistemáticas de los estudiantes que hacían que rindieran en la primera aplicación de manera diferente, hacen que a lo largo del tiempo su crecimiento sea constante.

Una vez que se han planteado subsituaciones donde se fija variabilidad en uno de los coeficientes (intercepto o pendiente) del nivel de estudiante, junto con el error intrasujeto, en la subsituación 6 se presentan modelos similares a los anteriores pero referidos al nivel de aula.

En los modelos de la **subsituación 6** se fija solo variabilidad en uno de los coeficientes aleatorios del nivel de aula (estatus o pendiente), y en el error intrasujeto del nivel del tiempo.

En el **modelo 6,** la varianza del error intrasujeto se fija a un 95% de la varianza total del modelo y, el resto de variabilidad se fija en la pendiente del nivel de aula. En este modelo se fijan los mismos porcentajes de varianza que en el modelo 3. Pero mientras que en el modelo 3, el 5% de varianza se fijaba en la pendiente del nivel de estudiante, en el modelo 6 se fija en la pendiente del nivel de aula.

En el **modelo 6.1** se fija el mismo porcentaje de varianza en el error intrasujeto que en el modelo 4, es decir, un 95%, el resto de varianza, que en el modelo 4 se fijaba en el estatus del nivel de estudiante, en el modelo 6.1 se fija en el estatus del nivel de aula.

Tabla V.8. *Modelo 6 y modelo 6.1*

Coeficientes	Valores de los coeficientes
Estatus o intercepto (β_{000})	**4,6**
Pendiente (β_{100})	**0,4**
Correlación entre el intercepto y la pendiente en el nivel de aulas (Raulas)	0
Correlación entre el intercepto y la pendiente en el nivel de estudiantes (Resudiantes)	0
Varianza del intercepto en el nivel de aulas (u_{00j})	**5%**
Varianza de la pendiente en el nivel de aulas (u_{10j})	**5%**
Varianza del intercepto en el nivel de estudiantes (r_{0ij})	0
Varianza de la pendiente en el nivel de estudiantes (r_{1ij})	0
Varianza del error intrasujeto en el nivel de tiempo (e_{tij})	**95%/95%**

El modelo 6**,** representa una situación que no se había estudiado hasta el momento, y, es que, ahora el rendimiento de los estudiantes a lo largo del tiempo se debe al efecto exclusivo de las características de las aulas, y no de los estudiantes. Todos los estudiantes, en general, parten de los mismos niveles de rendimiento y su crecimiento es constante (no se fija varianza ni en el intercepto ni en la pendiente del nivel de estudiante). Tan solo se fija variabilidad en la pendiente del nivel de aula, lo que supone que el rendimiento de partida de los estudiantes tanto de una misma aula como del resto de aulas es el mismo. Sin embargo, aunque las pendientes de crecimiento de los estudiantes de una misma aula son similares, las pendientes de crecimiento de los estudiantes que van a distintas aulas son diferentes, ya que se ha fijado varianza en la pendiente del nivel de aula.

Algo similar ocurre en el modelo 6.1, pero en vez de fijarse variabilidad en la pendiente del nivel de aula se fija en el estatus de ese mismo nivel. Esto supone que los estudiantes que van a una misma clase parten del mismo rendimiento previo, pero ese rendimiento inicial es diferente entre clases (no se fija varianza en el estatus del nivel de estudiante, pero si en el del nivel de aula). Sin embargo, las pendientes de crecimiento de los estudiantes son constantes independientemente de si van a la misma aula o a aulas distintas (no se fija varianza ni en la pendiente del nivel de estudiante ni del nivel de aula).

Posteriormente, se especifican una serie de modelos donde además de fijar varianza en el error intrasujeto, se fija simultáneamente variabilidad en el intercepto y la pendiente de un mismo nivel (subsituaciones 7 y 8).

En los modelos de la **subsituación 7** la suma del porcentaje de varianza fijado en el estatus y en la pendiente, no es superior a la varianza fijada en el error intrasujeto.

Concretamente, en el **modelo 7,** la suma de las varianzas del nivel de estudiante es un 5% de la varianza total del modelo; así, la varianza del intercepto del nivel de estudiante se fija a un 2,5% y, la varianza de la pendiente de ese mismo nivel al otro 2,5%. El resto de varianza se fija en el error intrasujeto (95% de varianza).

El porcentaje de varianza total del nivel de aula, estudiante y del error intrasujeto es el mismo que el que se fijó en el modelo 2 y modelo 4. Pero en el modelo 2, el 5% de varianza total del nivel de estudiante se fijó únicamente en la varianza de la pendiente, y en el modelo 4 en la varianza del estatus.

En cuanto al **modelo 7.1**, al igual que en los modelos 2.2 y 4.2, en el nivel de estudiante se fija un 50% de la varianza total del modelo y, en la varianza del error intrasujeto el otro 50%. Pero en el modelo 7.1, dentro del nivel de estudiantes ese 50% esta repartido en porcentajes iguales, entre la varianza de la pendiente del nivel de estudiante (25%) y, la varianza del intercepto de ese mismo nivel (25%).

En los modelos anteriores de esta subsituación, dentro del nivel de estudiante se ha fijado el mismo porcentaje de varianza en el intercepto que en la pendiente. Sin embargo, en el **modelo 7.2,** la varianza del estatus fijada es superior a la varianza de la pendiente, algo común en los modelos jerárquicos lineales utilizados en las evaluaciones educativas, tal y como se puede observar en las investigaciones de Bryk y Raudenbush (2002); Castro et al., (2008); Navarro (2013); Stevens y Zvoch (2006); Snijders y Bosker, (1999); Zvoch y Steve, (2006), etc.) En este modelo 7.2 la varianza de la pendiente del nivel de estudiante se fija a un 10% de la varianza total y, la del intercepto a un 40%, fijándose el resto de varianza en el error intrasujeto (50%).

A modo de ejemplo, los porcentajes de este modelo 7.2, son fijados en el **modelo 7.3** pero referidos al nivel de aula en vez de estudiante. De tal manera que la varianza del intercepto de dicho nivel se fija a un 40%, la varianza de la pendiente a un 10%, y el 50% de varianza restante, se fija en el error intrasujeto. Los resultados de dicho modelo pueden ser consultados en el Anexo III

Finalmente, dentro de esta subsituación, se desarrolla el **modelo 7.4,** donde el porcentaje de varianza de la pendiente del nivel dos, es superior (40%) a la varianza del intercepto de ese mismo nivel (10%), lo contrario a lo que ocurria en el modelo 7.2. El resto de varianza se fija en el error intrasujeto (50%).

Tabla V.9. *Modelo 7, modelo 7.1, modelo 7.2 y modelo 7.4*

Coeficientes	Valores de los coeficientes
Estatus o intercepto (β_{000})	**4,6**
Pendiente (β_{100})	**0,4**
Correlación entre el intercepto y la pendiente en el nivel de aulas (Raulas)	0
Correlación entre el intercepto y la pendiente en el nivel de estudiantes (Resudiantes)	0
Varianza del intercepto en el nivel de aulas (u_{00j})	0
Varianza de la pendiente en el nivel de aulas (u_{10j})	0
Varianza del intercepto en el nivel de estudiantes (r_{0ij})	**2,5%/17,5%/2,5%/40%/10%**
Varianza de la pendiente en el nivel de estudiantes (r_{1ij})	**2,5%/17,5%/2,5%/10%/40%**
Varianza del error intrasujeto en el nivel de tiempo (e_{tij})	**95%/65%/50%/50%/50%**

Los modelos anteriores, a excepción del modelo 7.3, muestran una situación donde en general, los estudiantes parten de distintos niveles de rendimiento previo y crecen a distinto ritmo, debido a las características sistemáticas de los mismos (se fija varianza en el intercepto y en la pendiente del nivel de estudiante). Sin embargo, dichos niveles de rendimiento no se ven afectados por las características de las aulas, ya que, no se fija variabilidad ni en el intercepto ni en la pendiente de dicho nivel. Lo contrario a lo que ocurre en el modelo 7.3, puesto que el rendimiento promedio inicial y el obtenido a lo largo del tiempo se debe a las características de las aulas y no de los estudiantes.

En este modelo 7.3, los estudiantes de una misma aula parten del mismo nivel de rendimiento y ese crecimiento es constante a lo largo del tiempo. Pero, el rendimiento, tanto en la primera ocasión como el resto de ocasiones de medida, es distinto entre las aulas, ya que las mismas cuentan con características diferentes que afectan de distinta manera al rendimiento de los estudiantes. Siendo dicho efecto el mismo sobre los niveles de logro de los estudiantes que van a la misma clase.

Si en los modelos de la subsituación 7 se fijan valores donde el porcentaje de varianza del error intrasujeto, es igual o superior a la varianza total del nivel de estudiante o del nivel de aula; en los modelos de la **subsituación 8**, el porcentaje de varianza total fijado en el nivel de estudiantes, es superior a la varianza del error intrasujeto.

En el **modelo 8,** la varianza total del nivel de estudiante suma un 95% de la varianza total del modelo (la varianza del intercepto se fija a un 47,5% de la varianza total del nivel de estudiante y la varianza de la pendiente a otro 47,5%), el 5% restante se fija en el error intrasujeto. Estos mismos porcentajes totales de varianza del nivel de estudiante, se fijaron en los modelos 3 y 5, pero la diferencia existente entre los tres modelos, es el reparto de dicho porcentaje de varianza dentro del nivel dos.

En el **modelo 8.1**, se fija el mismo porcentaje de varianza tanto en el error intrasujeto, como en la varianza de la pendiente y del intercepto del nivel de estudiante (33,3%). Las sumas de ambas varianzas del nivel de estudiante suponen un 66,6 % de la varianza total del modelo.

Finalmente, se desarrolla el **modelo 8.2** donde la varianza del error intrasujeto se fija a 0. Debiendose toda la varianza del modelo al nivel de estudiante (el 50% de la varianza total se fija en el intercepto y el otro 50% en la varianza de la pendiente). Este modelo en el cual no existe varianza en el error intrasujeto y, toda la varianza es fijada en el nivel dos, es similar a los modelos 3.2 y 5.2, ya que los porcentajes de varianza totales de cada uno de los niveles de anidamiento, son los mismos.

Tabla V.10. *Modelo 8, modelo 8.1 y modelo 8.2*

Coeficientes	Valores de los coeficientes
Estatus o intercepto (β_{000})	**4,6**
Pendiente (β_{100})	**0,4**
Correlación entre el intercepto y la pendiente en el nivel de aulas (Raulas)	0
Correlación entre el intercepto y la pendiente en el nivel de estudiantes (Resudiantes)	0
Varianza del intercepto en el nivel de aulas (u_{00j})	0
Varianza de la pendiente en el nivel de aulas (u_{10j})	0
Varianza del intercepto en el nivel de estudiantes (r_{0ij})	**47,5%/33,3%/50%**
Varianza de la pendiente en el nivel de estudiantes (r_{1ij})	**47,5%/33,3%/50%**
Varianza del error intrasujeto en el nivel de tiempo (e_{tij})	**5%/33,3%/0%**

En la mayoría de los modelos de la **subsituación 9** que se presentan a continuación, se fija variabilidad en la pendiente o en el intercepto de ambos niveles (estudiante y aula), junto con el error intrasujeto.

En el **modelo 9**, la varianza de la pendiente del nivel de aula se fija a un 2,5% de la varianza total del modelo, y la varianza de la pendiente del nivel de estudiante a otro 2,5%. El resto de parámetros de varianza de dichos niveles, es decir, el estatus del nivel de estudiante y de aula se fijan a 0, por lo tanto, la varianza restante, esto es el 95%, se fija en el error intrasujeto. Este modelo cuenta con los mismos porcentajes de varianza en la pendiente y en el error intrasujeto que el modelo 2.

Un modelo similar al modelo anterior, es el **modelo 9.1** cuyo resultado se adjunta en el Anexo III. En dicho modelo el porcentaje de varianza total, se encuentra repartido, entre la varianza del error intrasujeto (95%), la varianza del estatus del nivel de aula (2,5%) y la varianza de la pendiente del nivel de estudiante (2,5%).

También en el Anexo III, se presentan los resultados del **modelo 9.2** en el cual**,** la varianza del estatus del nivel de aula se fija a un 2,5% de la varianza total del modelo. Este mismo porcentaje de varianza se fija en el intercepto del nivel de estudiante, de esta manera la suma de ambas varianzas, de ambos niveles, (nivel 3 y

2) suponen un 5% de la varianza total del modelo. El 95% de la varianza restante se fija en el error intrasujeto.

Tabla V. 11. *Modelo 9*

Coeficientes	Valores de los coeficientes
Estatus o intercepto ($\boldsymbol{\beta_{000}}$)	**4,6**
Pendiente ($\boldsymbol{\beta_{100}}$)	**0,4**
Correlación entre el intercepto y la pendiente en el nivel de aulas (Raulas)	0
Correlación entre el intercepto y la pendiente en el nivel de estudiantes (Resudiantes)	0
Varianza del intercepto en el nivel de aulas ($\mathbf{u_{00j}}$)	0
Varianza de la pendiente en el nivel de aulas ($\mathbf{u_{10j}}$)	**2,5%**
Varianza del intercepto en el nivel de estudiantes ($\mathbf{r_{0ij}}$)	0
Varianza de la pendiente en el nivel de estudiantes ($\mathbf{r_{1ij}}$)	**2,5%**
Varianza del error intrasujeto en el nivel de tiempo ($\mathbf{e_{tij}}$)	**95%**

En el modelo 9, el rendimiento de los estudiantes a lo largo del tiempo, se debe tanto a las características de los estudiantes, como a las características de las aulas. Los estudiantes de una misma aula parten de un nivel de rendimiento previo que es similar, pero su rendimiento a lo largo del tiempo es diferente (no se fija varianza en el intercepto del estudiante, pero si en la pendiente). Dicho nivel de logro en las diferentes ocasiones de medida es distinto entre las aulas, pero, el nivel de rendimiento en la primera ocasión de medida es el mismo en todas las aulas. Se fija variabilidad en la pendiente del nivel de aula, pero no en el intercepto.

En el modelo 9.1, se presenta una situación en la cual los estudiantes parten de los mismos niveles de logro, pero su rendimiento a lo largo del tiempo es distinto debido a sus propias características. Además, en este modelo los estudiantes de una misma aula, tienen distinto rendimiento a lo largo del tiempo, pero parten de los mismos niveles de logro, siendo dicho logro diferente entre las aulas, al igual que la pendiente de crecimiento.

Mediante el modelo 9.2, se representa una situación en la cual todos los estudiantes crecen al mismo ritmo, pero sus niveles de logro de partida son distintos, debido tanto al efecto de sus propias características, como de las características de las aulas. Dentro de una misma aula, aunque los estudiantes cuentan con la misma pendiente de crecimiento, su rendimiento en la primera medida es diferente. Entre las aulas el rendimiento de partida es distinto, pero la pendiente de crecimiento es similar.

Con los modelos desarrollados hasta el momento se pretende estudiar, mediante el análisis del patrón de correlación obtenido en las puntuaciones de los estudiantes entre las distintas mediciones realizadas, la fiabilidad de las estimaciones en una serie de situaciones previas en las que solo se fija variabilidad en algunos de los coeficientes aleatorios de un modelo multinivel de carácter longitudinal (objetivo específico 2, *situación a*).

En los siguientes modelos que se van a explicar, desde la situación 10 a la 14, también tienen como objetivo estudiar la precisión en la estimación de las puntuaciones de los estudiantes. Pero ahora referido a una realidad en la que además de fijar varianza en el error intrasujeto y en el intercepto y la pendiente de un mismo nivel (aula o estudiante), se fija también distintos valores de correlación entre el intercepto y la pendiente de ese mismo nivel (objetivo específico 2**,** *situación b*).

- ***Modelos generados en la situación b***

La importancia de los modelos que se presentan a continuación, reside en estudiar no sólo cómo afectan los porcentajes de varianza fijados en cada coeficiente aleatorio, sino cómo la intensidad y la direccionalidad de los valores de correlación establecidos entre el intercepto y la pendiente, influyen en los patrones de correlación de las puntuaciones de los estudiantes, en las distintas ocasiones de medida.

En la mayoría de los modelos desarrollados en esta *situación b*, los valores de correlación entre el intercepto y la pendiente del nivel de estudiante o del nivel de aula, se fijan a un valor de correlación bajo positivo (0,2), alto positivo (0,9), bajo negativo (-0,2) o alto negativo (-0,9).

En los modelos de la **subsituación 10**, se fijan los mismos porcentajes de varianza que en el modelo 7, esto es, la varianza del error intrasujeto se fija a un 95% y la varianza del intercepto y de la pendiente del nivel de estudiante a un 2,5%.

Con estos porcentajes de varianza, en el **modelo 10** se fija una correlación entre el intercepto y la pendiente del nivel de estudiante de 0,2.

En el **modelo 10.1** el valor de la correlación fijado entre el intercepto y la pendiente del nivel de estudiante, es alto y positivo (0,9).

En los modelos 10 y 10.1 se han fijado valores de correlación con un sentido de relación positiva. Por ello, en los modelos 10.2 y 10.3 se fija una correlación negativa. Concretamente en el **modelo 10.2**, se fija una correlación de -0,2 mientras que en el **modelo 10.3** dicho valor de correlación se fija a -0.9.

Tabla V.12. *Modelo 10, modelo 10.1, modelo 10.2 y modelo 10.3*

Coeficientes	Valores de los coeficientes
Estatus o intercepto (β_{000})	**4,6**
Pendiente (β_{100})	**0,4**
Correlación entre el intercepto y la pendiente en el nivel de aulas (Raulas)	0
Correlación entre el intercepto y la pendiente en el nivel de estudiantes (Resudiantes)	**0,2/0,9/-0,2/-0,9**
Varianza del intercepto en el nivel de aulas (u_{00j})	0
Varianza de la pendiente en el nivel de aulas (u_{10j})	0
Varianza del intercepto en el nivel de estudiantes (r_{0ij})	**2,5%**
Varianza de la pendiente en el nivel de estudiantes (r_{1ij})	**2,5%**
Varianza del error intrasujeto en el nivel de tiempo (e_{tij})	**95%**

En los modelos anteriores de la subsituación 10, siempre se ha fijado el mismo porcentaje de varianza en el estatus que en la pendiente. A continuación, se desarrollan modelos donde la varianza del intercepto, no es la misma que la varianza de la pendiente.

En los modelos de la **subsituación 11,** se fijan los mismos porcentajes de varianza que los fijados en el modelo 7.3. Esto es, la varianza del intercepto del nivel de estudiante se fija un 10% de la varianza total, la varianza de la pendiente de ese mismo nivel a un 40%, y la varianza restante (50%) se fija en el error intrasujeto.

En el **modelo 11**, manteniendo los porcentajes de varianza fijados para todos los modelos de la subsituación 11, se fija una correlación entre el estatus y la pendiente del nivel de estudiante de 0,2. En el **modelo 11.1** dicha correlación entre el intercepto y la pendiente se fija a 0,9.

Un modelo similar al modelo 11.1, pero referido al nivel de aulas es el **modelo 11.2,** cuyos resultados aparecen en el Anexo III. En este modelo la varianza del intercepto del nivel de aula se fija a un 40% de la varianza total del modelo, la varianza de la pendiente de ese mismo nivel a un 10%, y el 50% restante se fija en el error intrasujeto. En cuanto a la correlación entre el intercepto y la pendiente del nivel de aula se fija a 0.9.

En el **modelo 11.3,** se fija la correlación entre el intercepto y la pendiente del nivel de estudiante a -0,2. Este mismo valor de correlación, pero referido a la relación entre el intercepto y la pendiente del nivel de aula, y, por lo tanto, fijando variabilidad en dicho nivel en vez de en el del estudiante, se corresponde con el **modelo 11.4,** cuyos resultados se presentan en el Anexo III. Concretamente, en dicho modelo, el porcentaje fijado en la varianza del intercepto del nivel de aula es de un 40%, el fijado en la pendiente de ese mismo nivel es de un 10%, y en cuanto a la varianza del error intrasujeto se fija a un 50%.

Finalmente, dentro de esta subsituación, en el **modelo 11.5**, se fija una correlación entre el intercepto y la pendiente de dicho nivel de -0,9.

Tabla V.13. *Modelo 11, modelo 11.1, modelo 11.3 y modelo 11.5*

Coeficientes	Valores de los coeficientes
Estatus o intercepto (β_{000})	**4,6**
Pendiente (β_{100})	**0,4**
Correlación entre el intercepto y la pendiente en el nivel de aulas (Raulas)	0
Correlación entre el intercepto y la pendiente en el nivel de estudiantes (Resudiantes)	**0,2/0,9/-0,2/-0,9**
Varianza del intercepto en el nivel de aulas (u_{00j})	0
Varianza de la pendiente en el nivel de aulas (u_{10j})	0
Varianza del intercepto en el nivel de estudiantes (r_{0ij})	**40%**
Varianza de la pendiente en el nivel de estudiantes (r_{1ij})	**10%**
Varianza del error intrasujeto en el nivel de tiempo (e_{tij})	**50%**

En los modelos desarrollados dentro de la **subsituación 12**, la varianza total del nivel de estudiante se fija a un 95%, fijándose la varianza del estatus de dicho nivel a un 47,5% y la varianza de la pendiente a otro 47,5%. En cuanto al error intrasujeto se fija a un 5% (mismos porcentajes de varianza que en el modelo 8).

Con estos valores de dispersión, en el **modelo 12**, se fija un valor de correlación entre el intercepto y la pendiente del nivel de estudiante de 0,2, en el **modelo 12.1** de 0,9, en el **modelo 12.2** de -0,2 y en el **modelo 12.3** de -0,9.

Tabla V.14. *Modelo 12, modelo 12.1 y modelo 12.3*

Coeficientes	Valores de los coeficientes
Estatus o intercepto (β_{000})	**4,6**
Pendiente (β_{100})	**0,4**
Correlación entre el intercepto y la pendiente en el nivel de aulas (Raulas)	0
Correlación entre el intercepto y la pendiente en el nivel de estudiantes (Resudiantes)	**0,2/0,9/-0,2/-0,9**
Varianza del intercepto en el nivel de aulas (u_{00j})	0
Varianza de la pendiente en el nivel de aulas (u_{10j})	0
Varianza del intercepto en el nivel de estudiantes (r_{0ij})	**47,5%**
Varianza de la pendiente en el nivel de estudiantes (r_{1ij})	**47,5%**
Varianza del error intrasujeto en el nivel de tiempo (e_{tij})	**5%**

Los siguientes modelos desarrollados, referidos a la **subsituación 13,** son modelos en los cuales se fija los mismos porcentajes de variabilidad. Así pues, la varianza de la pendiente del nivel de estudiante se fija a 33,3%, la varianza del estatus de ese mismo nivel a otro 33,3% y finalmente la varianza del error intrasujeto se fija también a un 33,3% (mismos porcentajes que en el modelo 8.1).

Con estos porcentajes de varianza, en el **modelo 13,** se fija un valor de correlación entre el intercepto y la pendiente del nivel de estudiante de 0,2, en el **modelo 13.1** de 0,9, en el **modelo 13.2** de -0,2 y en el **modelo 13.3** de -0,9.

Tabla V.15. *Modelo 13, modelo 13.1, modelo 13.2 y modelo 13.3*

Coeficientes	Valores de los coeficientes
Estatus o intercepto (β_{000})	**4,6**
Pendiente (β_{100})	**0,4**
Correlación entre el intercepto y la pendiente en el nivel de aulas (Raulas)	0
Correlación entre el intercepto y la pendiente en el nivel de estudiantes (Resudiantes)	**0,2/0,9/-0,2/-0,9**
Varianza del intercepto en el nivel de aulas (u_{00j})	0
Varianza de la pendiente en el nivel de aulas (u_{10j})	0
Varianza del intercepto en el nivel de estudiantes (r_{0ij})	**33,3%**
Varianza de la pendiente en el nivel de estudiantes (r_{1ij})	**33,3%**
Varianza del error intrasujeto en el nivel de tiempo (e_{tij})	**33,3%**

Finalmente, dentro de la *situación b*, se desarrollan cuatro modelos englobados en la **subsituación 14**. Estos modelos difieren en cierta medida de los modelos simulados hasta el momento dentro de la *situación b,* en que en los mismos se fija un cierto porcentaje de varianza en todos los coeficientes aleatorios menos en la varianza de la pendiente del nivel de estudiante. Por lo tanto, al no fijar variabilidad en dicho coeficiente, no es posible fijar valores de correlación entre el intercepto y la pendiente de dicho nivel dos.

Concretamente, en los modelos desarrollados en la situación 14, el porcentaje de varianza fijado en el intercepto del nivel de aula es de un 55%, el porcentaje de varianza fijado en la pendiente de ese mismo nivel es un 25%, el porcentaje de varianza del estatus del nivel de estudiante se fija a un 15% y, el 5% de varianza restante se fija en la varianza del error intrasujeto. Los cuatro modelos desarrollados dentro de esta situación, difieren entre sí en el valor de correlación fijado entre el intercepto y la pendiente del nivel de aula.

En el **modelo 14**, se fija la correlación a 0,2, en el **modelo 14.1** a 0,9, en el **modelo 14.2** a -0,2 y finalmente en el **modelo 14.3** a -0,9.

Tabla V.16. *Modelo 14, modelo 14.1, modelo 14.2 y modelo 14.3*

Coeficientes	Valores de los coeficientes
Estatus o intercepto (β_{000})	**4,6**
Pendiente (β_{100})	**0,4**
Correlación entre el intercepto y la pendiente en el nivel de aulas (Raulas)	**0,2/0,9/-0,2/-0,9**
Correlación entre el intercepto y la pendiente en el nivel de estudiantes (Resudiantes)	0
Varianza del intercepto en el nivel de aulas (u_{00j})	**55%**
Varianza de la pendiente en el nivel de aulas (u_{10j})	**25%**
Varianza del intercepto en el nivel de estudiantes (r_{0ij})	**15%**
Varianza de la pendiente en el nivel de estudiantes (r_{1ij})	0
Varianza del error intrasujeto en el nivel de tiempo (e_{tij})	**5%**

Los modelos de la subsituación 14, se desarrollan con el fin de conocer la fiabilidad de las estimaciones de las puntuaciones de los estudiantes a lo largo del tiempo, cuando dichas puntuaciones no se encuentran afectadas por las características sistemáticas de los estudiantes (no se fija variabilidad en la pendiente del nivel dos), tan sólo por las características de las aulas, junto con el efecto del valor de la correlación entre el intercepto y la pendiente de dicho nivel, y la varianza del error intrasujeto.

Tal y como se ha comentado en los párrafos precedentes, en los modelos desarrollados en la *situación b* (modelos de la subsituación 10 a la 14), se pretende estudiar no solo el efecto de los porcentajes de varianza fijados en los diferentes coeficientes, sino, también, el efecto del valor de correlación fijado entre el intercepto y la pendiente

En los modelos en los que se fija un valor de correlación positivo (0,2 o 0,9) entre el intercepto y la pendiente del nivel de estudiante (modelo 10, modelo 10.1, modelo 11, modelo 11.1, modelos 11.2, modelo 12, modelo 12.1, modelo 13 y modelo 13.1), en general, se representa una situación en la que los estudiantes que parten de niveles de rendimiento bajos, sus niveles de logro no se incrementan mucho a lo largo del tiempo, mientras que los que parten de niveles de rendimiento

alto, sus niveles de logro crecen mucho a lo largo de las diferentes ocasiones de medida.

Cuando no se fija variabilidad en el nivel de estudiantes, solo en la varianza del intercepto y la pendiente del nivel de aula, y una correlación positiva entre el intercepto y la pendiente de dicho nivel (modelo 11.2, modelo 14 y modelo 14.1). El rendimiento de los estudiantes que asisten a una misma aula es el mismo, pero las aulas cuentan con diferentes niveles de logro entre sí, ya que se fija variabilidad tanto en el intercepto como en la pendiente del nivel de aula. En estas situaciones, las aulas que en general parten de niveles bajos de rendimiento, su logro a lo largo del tiempo no se incrementa mucho. Mientras que las aulas que muestran en la primera medición altos niveles de rendimiento, los niveles de logro de las mismas se incrementan mucho a lo largo del resto de mediciones.

Cuando en el nivel de estudiante se fija una correlación negativa entre el intercepto y la pendiente, tal y como ocurre en los modelos 10.2, 10.3, 11.3, 11.5, 12.2, 12.3, 13.2 y 13.3. Los estudiantes que parten de bajos niveles de rendimiento académico, dicho nivel de logro a lo largo del tiempo crece mucho, mientras que los que parten de niveles de logro alto, su rendimiento no crecerá mucho a lo largo del tiempo o bien se mantiene constante. Aquí es posible que se de el denominado efecto techo, es decir, el crecimiento de los estudiantes que parten de un nivel de rendimiento alto, disminuye porque se aproximan a las puntuaciones máximas de la escala (Betebenner & Linn, 2009).

En esta situación, en la cual se fija una correlación negativa entre el intercepto y la pendiente del nivel de estudiante, se puede considerar que todo el sistema educativo es compensatorio, ya que se pretende equilibrar el rendimiento de los estudiantes con bajos niveles de logro inicial, con el de los estudiantes que parten de niveles altos de rendimiento.

Cuando se fija una correlación negativa entre el intercepto y la pendiente del nivel de aula y todos los estudiantes, en general, parten del mismo nivel de rendimiento previo, y su crecimiento académico es constante a lo largo del tiempo, no se fija varianza ni en el intercepto, ni en la pendiente del nivel de estudiante (modelo 11.4, modelo 14.2 y modelo 14.3), el rendimiento de los estudiantes que asisten a una misma aula es el mismo, pero debido al efecto de las aulas (se fija

varianza en el intercepto y en la pendiente del nivel de aula), estas difieren entre sí, tanto en los niveles iniciales promedios, como en los que alcanzan a lo largo del tiempo.

Al fijarse una correlación negativa entre el intercepto y la pendiente del nivel de aula, las aulas que estén compuestas en su mayoría por estudiantes con bajo rendimiento, el logro de las mismas crecerá mucho a lo largo del tiempo. Mientras que las que estén compuestas, sobre todo, por estudiantes con altos niveles de logro, mostrarán un rendimiento similar y/o que no se incrementa mucho a lo largo de las ocasiones de medida, pudiendo llegar a ocurrir nuevamente el denominado efecto techo, en el cual no se puede medir más a los estudiantes con niveles inciales de rendimiento altos. Fijar una correlación negativa entre el intercepto y la pendiente del nivel de aula, también representa una situación compensatoria, ya que las características de las aulas pretenden igualar, las diferencias en rendimiento entre aquellas aulas que parten de altos niveles de logro, y las que parten de bajos niveles de rendimiento.

Dichas tendencias explicadas anteriormente, serán mas o menos pronunciadas dependiendo de la fuerza de asociación de la correlación entre el intercepto y la pendiente. En los modelos en los que se fija una correlación baja de 0,2 (Wonnacott & Wonnacott, 1999), en general se observará una tendencia en la que el rendimiento del que parten los estudiantes no se encuentra muy relacionado con el que alcanzan a lo largo del tiempo. Sin embargo, cuando se fija una correlación de 0,9, al considerarse una correlación alta (Wonnacott & Wonnacott, 1999), el rendimiento del que parten los estudiantes está muy relacionado con el que alcanzan a lo largo del tiempo.

A la hora de interpretar el valor de correlación fijado entre el intercepto y la pendiente, es necesario tener en cuenta tanto el símbolo de la correlación, pues indica la direccionalidad de los niveles de logro, así como, la intensidad de la misma, ya que hace referencia la fuerza de relación entre los niveles de rendimiento de partida y los que se alcanzan a lo largo del tiempo.

Mediante estos primeros modelos iniciales que se han presentado hasta el momento, se pretende dar respuesta al objetivo específico 2, *situación a* y *b.* A partir de los mismos, se puede obtener una idea bastante clara de la influencia de los

porcentajes de varianza fijados en cada uno de los coeficientes y, del efecto de la correlación entre el intercepto y la pendiente, sobre los patrones de correlación de las puntuaciones de los estudiantes entre las distintas ocasiones de medida, y, por tanto, sobre la fiabilidad de las estimaciones en los modelos multinivel y longitudinal, así como del comportamiento del ERM.

Dichos modelos desarrollados hasta el momento, no suelen ser muy representativos de la realidad educativa, ya que generalmente en la misma, suele existir variabilidad en los distintos coeficientes aleatorios dentro de cada uno de los niveles de anidamiento. Muestra de ellos son los modelos que presentan, Bryk y Raudenbush (2002); Castro et al. (2008); Navarro (2013); Stevens y Zvoch (2006); Snijders y Bosker, (1999) y Zvoch y Steve, (2006).

Por ello, a continuación, siguiendo la lógica de los modelos presentados hasta el momento, se desarrollan una serie de modelos más complejos, pero asimismos más realistas. Estos modelos se enmarcan dentro del objetivo específico 2, *situación c*, ya que se fija variabilidad en todos los coeficientes aleatorios, esto es; en el error intrasujeto, el intercepto y la pendiente del nivel de estudiante y en el intercepto y la pendiente del nivel de aula. Además, se establecen distintos valores de correlación entre el intercepto y la pendiente, tanto del nivel 2 como del 3.

Como puede pensar el lector, debido a que se fija variabilidad en los coeficientes aleatorios de todos los niveles, junto con distintos valores de correlación, el número de modelos a desarrollar con diferentes combinaciones es muy elevado. Por ello se describen aquellos que se consideran que pueden ser más relevante para los estudiosos de la educación, ya que aportan información complementaria a la de los modelos desarrollados en la *situación a y b*.

- ***Modelos generados en la situación c***

Los modelos que se presentan en esta *situación c*, son modelos en los que el rendimiento de los estudiantes se debe tanto a las características sistemáticas de los mismos, como a las características de las aulas (se fija varianza en ambos niveles). A nivel de estudiante, en general, todos los estudiantes parten de distinto rendimiento previo (se fija varianza en el intercepto del nivel de estudiante) y su rendimiento a lo largo del tiempo también es distinto (se fija varianza en la pendiente de crecimiento del nivel de estudiante). A nivel de aula, en general, las aulas son

diferentes entre sí, tanto en el rendimiento inicial (se fija varianza en el estatus de dicho nivel tres) como en el nivel de logro que adquieren a lo largo del tiempo (se fija varianza en la pendiente del nivel de aulas). Además, se fijan distintos valores de correlación entre el intercepto y la pendiente del nivel de aula y del nivel de estudiante, de manera simultánea.

El primer modelo que se desarrolla dentro de esta *situación c,* es el **modelo 15**. En el mismo, los porcentajes de varianza se han especificado a partir de los resultados consultados en otras investigaciones (Castro, et al., 2009; Navarro; 2013; Zvoch & Stevens, 2003; Zvoch & Stevens, 2008), en las cuales se han trabajado con modelos semejantes a los que se vienen desarrollando. En estos modelos, el porcentaje de varianza total del nivel de estudiante, era superior al porcentaje de varianza total del nivel de aula, (por lo tanto, existía más variabilidad entre los estudiantes que entre las aulas). Dentro de cada uno de estos niveles, se observaba una mayor heterogeneidad en el rendimiento de partida que en los niveles de logro existentes a lo largo de las diferentes aplicaciones de medida, de ahí, que el porcentaje de la varianza del estatus fuera superior al porcentaje de la pendiente.

Teniendo en cuenta los resultados encontrados en las investigaciones citadas en el párrafo anterior, en el modelo 15 se fija la varianza del intercepto del nivel de aula a un 15% y, la varianza de la pendiente de ese mismo nivel a un 5%. En cuanto a los porcentajes de varianza del nivel de estudiante, el intercepto se fija a un 35% y la pendiente a un 25%. La varianza del error intrasujeto se fija un 20%, ya que en las diferentes investigaciones examinadas se encuentran entre el 20 y 25% de la varianza total del modelo (Castro et al., 2009; Navarro; 2013; Raudenbush & Bryk, 2002; Zvoch & Stevens, 2003; Zvoch & Stevens, 2008).

Finalmente, en cuanto a los valores de correlación fijados en dicho modelo, a nivel de aula se establece una correlación entre el intercepto y la pendiente de intensidad baja y de direccionalidad negativa (-0,15), valor que se ha tomado como referencia a partir de la investigación de Navarro (2013). A nivel de estudiante, la correlación fijada entre el estatus y la pendiente es también baja, pero de direccionalidad positiva (0,20), tal y como se muestra en las investigaciones de Castro et al. 2009; Gaviria y Castro, 2005; Navarro; 2013; Raudenbush y Bryk, 2002; Zvoch y Stevens, 2003; Zvoch y Stevens, 2008.

Tabla V.17. *Modelo 15*

Coeficientes	Valores de los coeficientes
Estatus o intercepto (β_{000})	**4,6**
Pendiente (β_{100})	**0,4**
Correlación entre el intercepto y la pendiente en el nivel de aulas (Raulas)	**-0,15**
Correlación entre el intercepto y la pendiente en el nivel de estudiantes (Resudiantes)	**0,20**
Varianza del intercepto en el nivel de aulas (u_{00j})	**15%**
Varianza de la pendiente en el nivel de aulas (u_{10j})	**5%**
Varianza del intercepto en el nivel de estudiantes (r_{0ij})	**35%**
Varianza de la pendiente en el nivel de estudiantes (r_{1ij})	**25%**
Varianza del error intrasujeto en el nivel de tiempo (e_{tij})	**20%**

En el siguiente modelo (**modelo 16**) también los valores de correlación fijados entre el intercepto y la pendiente del nivel de estudiante y aula, así como los porcentajes de varianza fijados en cada coeficiente aleatorio, han sido calculados a partir de un modelo real. Concretamente a partir de un modelo longitudinal (5 ocasiones de medida) y multinivel (tiempo, estudiante y escuela) presentado por Raudenbush y Bryk (2002)[14]

Para el desarrollo de dicho modelo 16, tomando como referencia los resultados obtenidos en el modelo de Raudenbush y Bryk (2002), se calculan los porcentajes de varianza total correspondiente a cada nivel, despejando la siguiente ecuación:

Ec.V.1

$$p = \frac{var(u_{00j}) + var\left(u_{10j}(T_{-1})\right) + 2cov\left(u_{00j} + u_{10j}(t_{-1})\right)}{var(e_{tij}) + var(r_{0ij}) + var\left(r_{1ij}(T_{-1})\right) + 2cov\left(r_{0ij} + r_{1ij}(t_{-1})\right) + var(u_{00j}) + var\left(u_{10j}(T_{-1})\right) + 2cov\left(u_{00j} + u_{10j}(t_{-1})\right)}$$

A modo de ejemplo en la ecuación V.1 se muestra el cálculo del porcentaje de varianza total del tercer nivel (aula en los modelos simulados hasta el momento y, escuela en el modelo de Raudenbush y Bryk, (2002)).

[14] Para la obtención de más información sobre el modelo consultar Raudenbush y Bryk (2002) consultar el libro *Hierarchical Lienal Models* (2002, p.240)

En el numerador de la ecuación anterior, se encuentra la suma de la varianza del intercepto del tercer nivel$\left(u_{00j}\right)$, más la varianza de la pendiente de ese mismo nivel multiplicada por la ocasión de medida correspondiente menos 1, $\left(u_{10j}(T_{-1})\right)$ (en este modelo se considera la última ocasión de medida), más dos veces la covarianza del intercepto y la pendiente, este último coeficiente se multiplica nuevamente por la ocasión de medida correspondiente, menos 1, $\left(2cov\left(u_{00j} + u_{10j}\,(t_{-1})\right)\right)$.

En el denominador se observa la suma de la varianza total del modelo, es decir, la varianza del error intrasujeto, $\left(e_{tij}\right)$, la varianza total del segundo nivel $\left(var\left(r_{0ij}\right) + var\left(r_{1ij}(T_{-1})\right) + 2cov\left(r_{0ij} + r_{1ij}\,(t_{-1})\right)\right)$, y finalmente la varianza total del tercer nivel $(var\left(u_{00j}\right) + var\left(u_{10j}(T_{-1})\right) + 2cov\left(u_{00j} + u_{10j}\,(t_{-1})\right))$

Si se despeja la fórmula anterior para cada nivel, con los valores obtenidos en el modelo multinivel de Raudenbush y Bryk, (2002), y considerando la última ocasión de medida de su modelo (ocasión 5), se obtiene una varianza total en el error intrasujeto de un 20%, en el nivel de estudiante (nivel 2) de un 59% y en el nivel de escuela (nivel 3) de un 21%.

A partir de los porcentajes de varianza totales obtenidos en cada uno de los niveles, dentro del segundo y tercer nivel, se calcula el porcentaje de varianza correspondiente a la varianza del estatus y de la pendiente

En la ecuación siguiente, a modo de ejemplo, se muestra el cálculo del porcentaje de varianza de la pendiente correspondiente al nivel dos.

$$p = \frac{var(r_{10j})}{var(r_{0ij}) + var(r_{1ij}(t_{-1})) + (2cov\left(r_{0ij} + r_{1ij}\,(t_{-1})\right)) * t_{-1}} \qquad \textbf{Ec.V.2}$$

En la ecuación anterior, en el numerador se encuentra la varianza de la pendiente del segundo nivel $\left(r_{1ij}\right)$ y, en el denominador la varianza total de dicho nivel multiplicado por la ocasión correspondiente menos 1.

Para el cálculo de la varianza del intercepto del nivel de estudiante, la fórmula es similar a la anterior, con la única diferencia de que ahora el numerador hace referencia a la varianza del intercepto y el denominador a la varianza total, sin que

la misma se encuentre multiplicada por el tiempo, tal y como se puede observar en la siguiente ecuación:

$$p = \frac{var(r_{0ij})}{var(r_{0ij}) + var(r_{1ij(t-1)}) + 2cov\left(r_{0ij} + r_{1ij(t-1)}\right)} \qquad \textbf{Ec.V.3}$$

Al resolver las fórmulas anteriores, a partir de los resultados del modelo de Raudenbush y Bryk, (2002), se obtienen los siguientes porcentajes de varianza de intercepto y pendiente.

En el nivel de estudiantes el porcentaje de varianza del intercepto obtenido, es un 64% sobre el 100% de la varianza de dicho nivel, pero sobre el 59%, que es la varianza total correspondiente al nivel de estudiante, es un 37,76%. La varianza de la pendiente obtenida en dicho nivel es un 11% sobre el 100% de varianza, lo que corresponde a un 6,49% sobre el 59% de la varianza total de dicho nivel, y finalmente, la covarianza es un 25% sobre 100% y un 14,75% sobre el 59%, que es el porcentaje correspondiente a la covarianza.

En el tercer nivel, recuérdese que la varianza total obtenida era un 21% de la varianza total del modelo. Así, aunque sobre el 100%, la varianza del intercepto calculada es de un 16%, sobre ese 21% es un 3,36%. La varianza de la pendiente obtenida es un 45% sobre 100% y, un 9,45% sobre el 21%. Finalmente, respecto a la covarianza obtenida es un 39% sobre 100% y un 8,19% sobre el 21%. En el error intrasujeto la varianza es el 100% sobre el 100% lo que supone un 20% de la varianza total del modelo.

Como en los modelos simulados no se fijan valores de covarianza entre el intercepto y la pendiente, sino que se simulan valores de correlación, despejando la fórmula siguiente se obtiene el valor de la misma, tanto para el nivel de estudiante, como para el nivel de aula.

$$r_{xy} = \frac{cov(\mathrm{u}_{00j}, \mathrm{u}_{10j})}{\sqrt{\mathrm{u}_{00j} * \mathrm{u}_{10j}}} \qquad \textbf{Ec.V.4}$$

En la expresión anterior (Ec.V.4) , donde a modo de ejemplo se muestra el cálculo de la correlación para el nivel de aula, en el numerador se encuentra el valor de la covarianza entre el intercepto y la pendiente y en el denominador la raíz de la varianza del intercepto (u_{00j}), que se multiplica por la raíz de la varianza de la pendiente (u_{10j}). Despejando esta ecuación, tanto para el nivel dos, como para el

nivel tres, se obtiene una correlación entre el intercepto y la pendiente del nivel de estudiante de 0,24 y, de 0,36 en el nivel de aula.

Al no fijarse valores de covarianza sino de correlación, en dicho modelo 16, las sumas de los porcentajes de varianza fijados no computan en el 100%, si no el 77%, ya que el 23% faltante se debe a la covarianza. Por ello, en la siguiente tabla, con el fin de facilitar la interpretación del modelo y seguir la lógica que se viene utilizando hasta el momento, donde las sumas de las varianzas corresponden al 100% de la varianza total del modelo, al lado de los porcentajes reales del modelo de Raudenbush y Bryk, (2002), aparece entre paréntesis el porcentaje correspondiente al considerar que todas las varianzas suman el 100%.

Tabla V.18. *Modelo 16*

Coeficientes	Valores de los coeficientes
Estatus o intercepto (β_{000})	**4,6**
Pendiente (β_{100})	**0,4**
Correlación entre el intercepto y la pendiente en el nivel de aulas (Raulas)	**0,36**
Correlación entre el intercepto y la pendiente en el nivel de estudiantes (Resudiantes)	**0,24**
Varianza del intercepto en el nivel de aulas (u_{00j})	**3,36% (4,35%)**
Varianza de la pendiente en el nivel de aulas (u_{10j})	**9,45% (12,27%)**
Varianza del intercepto en el nivel de estudiantes (r_{0ij})	**37,7% (48, 96%)**
Varianza de la pendiente en el nivel de estudiantes (r_{1ij})	**6,49%% (8,42%)**
Varianza del error intrasujeto en el nivel de tiempo (e_{tij})	**20% (26%)**

En los modelos siguientes de la **subsituación 17**, se continúa utilizando los porcentajes de varianza calculados a partir del modelo de Raudenbush y Bryk, (2002) en la ocasión 5, pero considerando que la suma de la varianza total es un 100%. Con esos porcentajes se fijan diferentes valores de correlación entre el intercepto y la pendiente del nivel dos y tres (0,2, 0,9, -0,2, -0,2).

Como se puede observar en la siguiente tabla (Tabla V.19) en los modelos de esta subsituación se fijan los siguientes porcentajes de varianza: en el tercer nivel, el porcentaje de varianza del intercepto se fija a un 4,35% y la varianza de la

pendiente a un 12,27%, la suma de ambas varianzas supone un 16,62% de la varianza total del modelo. En el nivel de estudiante, la varianza del intercepto se fija a un 48,96% y la de la pendiente a un 8,42%, por lo tanto, sumando ambas varianzas, en dicho nivel se encuentra el 57,38% de la varianza total del modelo, finalmente, la varianza restante se fija en el error intrasujeto, es decir, un 26%.

Tabla V.19. *Valores de los modelos de la subsituación 17*

Coeficientes	Valores de los coeficientes
Estatus o intercepto (β_{000})	**4,6**
Pendiente (β_{100})	**0,4**
Varianza del intercepto en el nivel de aulas (u_{00j})	**4,35%**
Varianza de la pendiente en el nivel de aulas (u_{10j})	**12,27%**
Varianza del intercepto en el nivel de estudiantes (r_{0ij})	**48,96%**
Varianza de la pendiente en el nivel de estudiantes (r_{1ij})	**8,42%**
Varianza del error intrasujeto en el nivel de tiempo (e_{tij})	**26%**

En cuanto a los valores de correlación fijados entre el intercepto y la pendiente, tanto del nivel de estudiante como del nivel de aula, son los mismos que los establecidos en casi todos los modelos de la *situación b*. Se fijan dos valores positivos (0,2 y 0,9) y dos valores negativos (-0,2 y -0,9) que representan una fuerza de intensidad alta (0,9) y baja (0,2) (Wonnacott & Wonnacott, 1999).

De esta manera, tal y como se puede observar en la siguiente tabla (Tabla V.20), dieciséis son las combinaciones posibles que dan como resultado a dieciséis modelos distintos dentro de esta subsituación 17.

Tabla V.20. *Combinaciones de correlación entre el estatus y la pendiente*

Combinaciones	R_{aulas}	$R_{estudiantes}$
1	0,2	0,2
2	0,2	0,9
3	0,2	-0,2
4	0,2	-0,9
5	0,9	0,2
6	0,9	0,9
7	0,9	-0,2
8	0,9	-0,9
9	-0,2	0,2
10	-0,2	0,9
11	-0,2	-0,2
12	-0,2	-0,9
13	-0,9	0,2
14	-0,9	0,9
15	-0,9	-0,2
16	-0,9	-0,9

En los modelos de la **subsituación 18,** la suma de las varianzas del nivel de aula y del nivel de estudiante, es un 5% de la varianza total del modelo, por ello cada coeficiente aleatorio se fija a 1,25%. El 95% restante se fija en la varianza del error intrasujeto.

Con estos porcentajes de varianza se desarrollan 16 modelos, que hacen referencia a cada una de las 16 combinaciones explicadas en la tabla anterior (Tabla V.20).

Tabla V. 21. *Valores de los modelos de la subsituación 18*

Coeficientes	Valores de los coeficientes
Estatus o intercepto (β_{000})	**4,6**
Pendiente (β_{100})	**0,4**
Varianza del intercepto en el nivel de aulas (u_{00j})	**1,25%**
Varianza de la pendiente en el nivel de aulas (u_{10j})	**1,25%**
Varianza del intercepto en el nivel de estudiantes (r_{0ij})	**1,25%**
Varianza de la pendiente en el nivel de estudiantes (r_{1ij})	**1,25%**
Varianza del error intrasujeto en el nivel de tiempo (e_{tij})	**95%**

En la **subsituación 19,** aunque también se desarrollan 16 modelos, referidos a cada una de las combinaciones explicadas en la Tabla V.20. Con el fin de sintetizar los resultados de los mismos, tan sólo se presentan la matriz de correlaciones de aquellas combinaciones que se consideran que pueden tener un mayor interés. Concretamente se presentan los resultados de la combinación 1, 2, 6, 9, 10, 11,13, 14, 15 y 16. Los resultados del resto de combinaciones pueden ser consultadas en el Anexo III.

En los modelos de esta subsituación 19, la varianza del error intrasujeto se fija a un 5% de la varianza total del modelo. El porcentaje total de varianza del nivel de estudiante a un 47,5% y el del nivel de aula a otro 47,5%. Dentro de cada uno de estos niveles, los coeficientes aleatorios (la varianza del intercepto y la pendiente) se fijan a un 23,75%.

Tabla V.22. *Valores de los modelos de la subsituación 19*

Coeficientes	Valores de los coeficientes
Estatus o intercepto (β_{000})	**4,6**
Pendiente (β_{100})	**0,4**
Varianza del intercepto en el nivel de aulas (u_{00j})	**23,75%**
Varianza de la pendiente en el nivel de aulas (u_{10j})	**23,75%**
Varianza del intercepto en el nivel de estudiantes (r_{0ij})	**23,75%**
Varianza de la pendiente en el nivel de estudiantes (r_{1ij})	**23,75%**
Varianza del error intrasujeto en el nivel de tiempo (e_{tij})	**5%**

En cuanto a la interpretación de los valores de correlación de estos modelos de la *situación c*, cuando en un modelo se fijan valores de correlación positivos entre el intercepto y la pendiente, tanto en el nivel de estudiante como en el nivel de aula, indica que en el nivel dos, los estudiantes que parten de niveles de rendimiento alto crecen mucho a lo largo del tiempo, mientras que los que parten de niveles de logro bajo, su rendimiento no se incrementa mucho a lo largo de las ocasiones de medida o se mantiene constante. En el nivel tres, las aulas que cuentan con altos niveles de rendimiento incial son las que más crecen, y las que parten de niveles de rendimiento más bajos, son las que mostrarán niveles de logro inferiores a lo largo del resto de ocasiones de medida. En estas condiciones en las cuales se fija una correlación positiva entre el intercepto y la pendiente, puede producirse el conocido efecto mateo, según el cual, los estudiantes o aulas con puntuaciones iniciales altas tenderán a crecer más.

En los modelos donde se fija una correlación negativa entre el intercepto y la pendiente del nivel de aula y del nivel de estudiante, los estudiantes que parten de niveles de rendimiento alto, en general no experimentan un alto incremento de sus niveles de rasgo a lo largo del tiempo, mientras que, por el contrario, los estudiantes que parten de bajos niveles de logro, su rendimiento crece mucho a lo largo de las diferentes ocasiones de medida; de esta manera puede existir un efecto compensatorio debido a las características del sistema educativo. A nivel de aulas, también se da dicho efecto compensatorio, ya que las aulas que parten, en general, de niveles de rendimiento bajo, su logro a lo largo del tiempo se incrementa mucho, mientras que las aulas que parten de niveles iniciales de logro que son elevados, el rendimiento de las mismas se mantiene constante o no crece mucho (bajo esta condición en la cual se fija una correlación negativa en el nivel de aula y de estudiante es posible que se de el efecto techo)

Cuando se fija una correlación positiva entre el intercepto y la pendiente del nivel de aula y una correlación negativa entre el intercepto y la pendiente del nivel de estudiante, se produce una situación compensatoria debido al efecto de las aulas, ya que dentro de una misma aula los estudiantes que parten de bajos niveles de rendimiento, crecen mucho a lo largo de las mediciones. Mientras que los que parten de altos niveles de logro se mantienen estable o no crecen mucho (es posible la

aparición del efecto techo). En cuanto al nivel tres, las aulas que cuentan con las medias mas altas alcanzan niveles de rendimiento muy elevados a lo largo del tiempo, que puede ser ocasionado por el efecto techo de la prueba o del aula, mientras que las que parten de niveles bajo de logro, el rendimiento de las mismas no se incrementa mucho a lo largo del tiempo.

Finalmente, si se fija una correlación negativa entre el intercepto y la pendiente del nivel de aula y una correlación positiva entre el intercepto y la pendiente del nivel de estudiante, las aulas que parten de niveles de rendimiento bajo, potencian a sus estudiantes para que alcancen niveles de rendimiento altos a lo largo del tiempo. Mientras que las que parten ya de un rendimiento alto, su rendimiento no se incrementa mucho a lo largo de las diversas aplicaciones (posible efecto techo). A nivel de estudiante, los que parten de un bajo nivel de logro, su rendimiento no se incrementa mucho a lo largo de las diferentes ocasiones de medida, sin embargo, los que parten de niveles de logro alto, su rendimiento si crece mucho a lo largo del tiempo (aquí también es posible el efecto mateo).

Estas tendencias serán más o menos pronunciadas en cada uno de los niveles, dependiendo de la intensidad de la relación fijada entre el intercepto y la pendiente del nivel de estudiante y del nivel de aula. Cuanto mayor sea el valor de la correlación más marcadas serán las mismas.

Una vez simulado los modelos explicados anteriormente, como conclusión final a este primer problema de investigación, se decidió dar respuesta a la siguiente pregunta: **¿Qué aulas potencia más el rendimiento de los estudiantes, haciendo que tengan una mayor tasa de crecimiento entre la primera ocasión de medida y la última?**

Para dar respuesta a esta pregunta, se utilizaron los datos generados en los cuatro modelos de la **subsituación 14**. La selección de la misma se debe a que, si se recuerda, son modelos donde no se fija variabilidad en la pendiente del nivel de estudiante, por lo tanto, el rendimiento académico de los mismos no se encuentra afectado por sus características sistemáticas, sino por las del aula. Además, en los modelos de esta subsituación se fijan diferentes valores de correlación entre el intercepto y la pendiente del nivel de aula (0,2, 0,9, -0,2 y -0,9), lo que también permite estudiar cómo afecta la relación entre el rendimiento de partida y el

obtenido a lo largo del tiempo en las aulas, en la tasa de crecimiento de los estudiantes.

Recuérdese que en los modelos de esta subsituación, los porcentajes de varianza fijados fueron: un 55% de varianza en el intercepto del nivel de aula, un 25% en la varianza de la pendiente de ese mismo nivel, un 15% en la varianza del estatus del nivel de estudiante, y un 5% en la varianza del error intrasujeto.

Para el estudio de las aulas que más potencian el rendimiento de los estudiantes, además de utilizar los resultados obtenidos con el programa R-studio, se utiliza como software estadístico el programa SPSS versión 23, en el cual se genera una base de datos compuesta por las 400 aulas, los 10.000 estudiantes agrupados en su correspondiente aula, los residuos de los distintos coeficientes aleatorios del modelo (u_{00j}, u_{10j}, r_{0ij}, y e_{tij}), (no se especifica el coeficiente aleatorio referido a la pendiente del nivel de estudiante, ya que su varianza se fijó a 0), y las puntuaciones promedias de los estudiantes en la primera ocasión de medida (Y_{0ij}) y en la última (Y_{8ij}). Se generan cuatro bases de datos referidas a cada uno de los cuatro modelos (modelo 14, modelo 14.1, modelo 14.2 y modelo 14.3) de la subsituación 14, en las cuales el procedimiento seguido es siempre el mismo, estas bases de datos pueden ser consultadas en el Anexo IV.

En dichas bases de datos se tipifican los residuos del estatus y la pendiente del nivel de aula (Zu_{00j}, Zu_{10j}), así como el residuo del estatus del nivel de estudiante (Zr_{0ij}), con la finalidad de facilitar su interpretación. Asimismo, se agrupan en quintiles a todas las aulas utilizando como variable de agrupación el residuo tipificado del intercepto del nivel de estudiante (Zr_{0ij}).

Una vez agrupadas todas las aulas en quintiles, se realizan descriptivos para conocer dentro de cada uno de los mismos cual es el valor máximo, mínimo y la media de Zr_{0ij}, y, así, poder seleccionar las aulas más cercanas a dichos valores.

Debido a que dentro de cada quintil puede haber una gran cantidad de aulas (recuérdese que el total de aulas generadas eran 400), se utiliza como criterios de selección: que sean aulas que tengan un valor de Zu_{00j}, cercano a la media obtenida en los descriptivos de Zr_{0ij} en cada uno de los quintiles, y que además se opten por aulas con tres tipos de pendiente, baja, media y alta, de acuerdo a su valor en Zr_{0ij}.

De esta manera se seleccionan un total de 15 aulas (3 aulas referidas a los tres tipos de pendiente: media, alta y baja, en cada uno de los 5 quintiles). Cada una de estas aulas están compuestas por estudiantes que se encuentran agrupados en los diferentes quintiles.

Una vez seleccionadas las aulas correspondientes a los tres tipos de pendiente en cada uno de los quintiles, con la base de datos completa se segmenta por aula y por la variable de agrupación ($\mathrm{Zr_{0ij}}$) creando una base final, con los resultados promedios de cada una de las aulas seleccionadas para cada uno de los cinco quintiles. En dicha base de datos se especifica el identificador del aula y, dentro de cada aula, los quintiles en los que se agrupan los estudiantes ($\mathrm{Z(r_{0ij})}\_q$), el número de estudiantes por quintil (N), la puntuación promedia de los estudiantes agrupados en los cinco quintiles en la primera ocasión de medida (Y_{0ij}) y en la última (Y_{8ij}), la diferencia entre la puntuación promedia obtenida entre la última y la primera ocasión de medida ($Y_{8ij} - Y_{0ij}$), los coeficientes aleatorios del nivel de aula y estudiante tipificados ($Z\mathrm{u_{00j}}$, $Z\mathrm{u_{10j}}$, $\mathrm{Zr_{0ij}}$) y, sin tipificar $\mathrm{u_{00j}}$, $(\mathrm{u_{00j}}, \mathrm{u_{10j}}, \mathrm{r_{0ij}})$, y los residuos del error intrasujeto en la primera y en la última ocasión de medida ($\mathrm{e_{0ij}}, \mathrm{e_{8ij}}$). Estos valores son los que se presentan en el capítulo de resultados para cada uno de los modelos de la subsituación 14, y los que se resumen de manera más visual en la siguiente tabla (Tabla V.23)

Tabla V.23. *Interpretación de los estadísticos para conocer el crecimiento de las aulas*

Estadístico	Interpretación
$\mathbf{Z(r_{0ij})\_q)}$	Quintil en el cual se agrupan los estudiantes acorde al valor del residuo tipificado de su intercepto
N	Número de estudiantes que hay en dicho quintil
$\boldsymbol{Y_{0ij}}$	Puntuación promedia de los estudiantes agrupados en un determinado quintil en la primera ocasión de medida
$\boldsymbol{Y_{8ij}}$	Puntuación promedia de los estudiantes agrupados en un determinado quintil en la última ocasión de medida (novena medición)
$\boldsymbol{(Y_{8ij} - Y_{0ij})}$	Diferencia entre la puntuación promedia obtenida entre la última ocasión de medida y la primera
$\boldsymbol{Zu_{00j}}$	Residuo del estatus del nivel de aula tipificado
$\boldsymbol{Zu_{10j}}$	Residuo de la pendiente del nivel de aula tipificado
$\mathbf{Zr_{0ij}}$	Residuo del estatus del nivel de estudiante tipificado
$\mathbf{u_{00j}}$	Residuo del estatus del nivel de aula sin tipificar
$\mathbf{u_{10j}}$	Residuo de la pendiente del nivel de aula sin tipificar
$\mathbf{r_{0ij}}$	Residuo del estatus del nivel de estudiante sin tipificar
$\mathbf{e_{0ij}}$	Residuo del error intrasujeto en la primera ocasión de medida
$\mathbf{e_{8ij}}$	Residuo del error intrasujeto en la última ocasión de medida

A partir de los datos de esta base, se crean las tablas de los resultados finales, cinco tablas referidas a los cinco quintiles, correspondientes a los cuatro modelos de la subsituación 14, lo que supone un total de 20 tablas. Cada una de ellas están compuestas por los datos promedios referidos a las 15 aulas seleccionadas de acuerdo a los criterios preestablecidos. Estas tablas se acompañan de gráficos elaborados en Excel, que permiten analizar de manera más visual en qué modelos

los estudiantes incrementan más su rendimiento entre la primera ocasión de medida y la última.

V.3.3.3.2. Generación de modelos del segundo problema de investigación

Para dar respuesta al segundo problema de investigación, a través del cual, se pretende conocer cómo afecta el tamaño de las aulas en la estimación de su pendiente media de crecimiento, se desarrollan una serie de modelos, acorde a cada una de las tres situaciones planteadas dentro del objetivo específico 2 de dicho problema.

En la generación de las diferentes condiciones muestrales fijadas en las *situaciones a, b y c*, se utilizan modelos donde los porcentajes de varianzas fijados en los coeficientes aleatorios, así, como los valores de correlación entre el intercepto y la pendiente del nivel de estudiante y de aula, son los establecidos en el modelo 15. Se utiliza dicho modelo como referencia, ya que sus porcentajes fueron fijados a partir de investigaciones basadas en la realidad educativa.

Recuérdese que en el **modelo 15**, el porcentaje de varianza del intercepto del nivel de aula se fijó a un 15% y el de la pendiente a un 5%. En el nivel de estudiante, la varianza del estatus se fijó a un 35% y la pendiente a un 25% y, finalmente, la varianza del error intrasujeto a un 20%. En cuanto a los valores de correlación entre el intercepto y la pendiente, en el nivel de aula se fijó a -0,15 y, en el nivel de estudiante a 0,20. Manteniendose dichos valores, sólo se modifica el porcentaje de aulas con diferente número de estudiantes (aulas pequeñas, medianas y grandes).

- ***Modelo de la situación a***

Para dar respuesta al segundo objetivo específico, del segundo problema de investigación (*ver apartado V.2 Objetivos de investigación*), en la *situación a* se simula un modelo (**modelo 15.a**) donde se pretende que alrededor de un 75% del total de aulas sean pequeñas, consideradas como tal, porque deben estar formadas por un mínimo de 5 estudiantes y un máximo de 10. Además, se pretende que un 12,5% de la muestra total de aulas, sean medianas, compuestas por un mínimo de 11 estudiantes y un máximo de 19, y, finalmente, que el otro 12,5% de las aulas, sean grandes, ya que deben estar formadas por un máximo de 30 estudiantes y un mínimo de 20.

- ***Modelo de la situación b***

En el modelo 15 de la *situación b* **(modelo 15.b)** se quiere que alrededor de un 75% de la muestra total de aulas, sean medianas, ya que, están compuestas por un mínimo de 11 estudiantes y un máximo de 19. Se han denominado medianas porque a partir de 20 estudiantes por unidad en el nivel macro, ya se considera un tamaño adecuado para poder realizar multinivel (Raudenbush & Bryk, 2002). En cuanto al 25% de aulas restantes, se pretende que estén formadas, alrededor de un 12,5% por aulas pequeñas, y entorno al otro 12,5% por aulas grandes.

- ***Modelo de la situación c***

Finalmente, en el **modelo15.c** (*situación c,* modelo 15), se quiere simular un tamaño muestral donde, en torno a un 75% de aulas sean grandes, ya que están formadas por más de 20 estudiantes. En cuanto al resto de las aulas, se pretende que alrededor de un 12,5% sean pequeñas y el otro 12,5% sean aulas medianas.

Una vez simulado los tres modelos de la *situación a, b y c*, se pretende analizar el porcentaje de aulas que se diferencian de la media respecto al porcentaje que no, prestando atención al tipo de aula (pequeña, mediana y grande). En definitiva, se quiere comprobar si la recomendación de Raudenbush y Bryk (2002), según la cual es necesario trabajar con 20 estudiantes por aula (aulas grandes) en los diseños multinivel, es adecuada para poder estimar de manera fiable la pendiente de crecimiento de las aulas.

Para ello, a partir de los resultados obtenidos en los modelos simulados en R-Studio (modelo 15.a, modelo 15.b y modelo 15.c), se genera una base de datos en SPSS, de cada uno de ellos. Cada base de datos esta compuesta por las aulas, los 10.000 estudiantes agrupados en su correspondiente aula, y los residuos de los distintos coeficientes aleatorios del modelo (u_{00j}, u_{10j}, r_{0ij}, r_{1ij}y e_{tij}).

Con estos datos, lo primero que se cálcula es el número total de estudiantes que hay en cada aula (N_estudiantes), con el fin de identificar el tamaño de cada una de ellas. Por ello, dicha variable es recodificada en una nueva (Tamaño_aulas), donde el valor 1 se refiere a aulas pequeñas, el 2 a aulas medianas y el 3 a aulas grandes.

Con el fin de conocer que aulas se diferencian de la media y cuales no, es necesario saber los intervalos de confianza, y clasificar a las mismas acorde a dichos valores. Para el cálculo de los intervalos de confianza hay que estimar el error muestral máximo(EMM), que requiere a su vez del cálculo del Error Típico (ET).

El ET se cálcula siguiendo la fórmula siguiente:

$$ET = \frac{sd_{(r_{1ij})}}{\sqrt{N-1}} \qquad \textbf{Ec.V.5}$$

Como se observa en la ecuación anterior V.5, para el cálculo del ET lo primero que se necesita es la desviación típica de la pendiente de crecimiento de los estudiantes, dicha desviación se divide entre la raíz del número de estudiantes menos 1.

Una vez cálculado el ET, es posible calcular el error muestral máximo (EMM) con un nivel de confianza del 95%, tal y como se muestra a continuación.

$$EMM = 1{,}96 * ET \qquad \textbf{Ec.V.6}$$

Conociendo el valor del EMM, se aplica la sintaxis del SPSS que se adjunta en el Anexo IV para el cálculo de los intervalos de confianza. Dicha sintaxis, genera una variable denominada R_cat_M1, que permite identificar a las aulas que no se diferencian de la media (tienen el valor de 0), las que están por debajo de la media (valor 1) y las que están por encima de la media (valor 2). Como realmente el interés reside en conocer las aulas que no se diferencian de la media (categoría 0), y las que se diferencian de la media (categoría 1 y 2), sin especificar si están por encima o por debajo de la misma, la categoría 2 se recodifica en la misma variable por la categoría 1.

Con estas bases de datos generadas para cada uno de los tres modelos (se adjuntan en el Anexo IV), ya es posible crear tablas de contingencia donde se puede estudiar el porcentaje de aulas que se diferencia de la media y el que no, dentro de cada una de las situaciones muestrales. El problema es que puede ocurrir que muchas aulas no se diferencien de la media no por el error de estimación de su pendiente de crecimiento, sino, porque, realmente cuentan con valores en su pendiente que hacen que se encuentren en la media.

Para distinguir, dentro de lo posible, entre estos dos tipos de aulas, las que se sitúan en la media debido a su error de estimación, y las que se situán en la media porque es donde les corresponde, previo a la realización de las tablas de contingencia se eliminan aquellas aulas cuyo residuo de crecimiento supone un 10% de su desviación típica. Concretamente, como en estos modelos la desviación típica de la pendiente de crecimiento del nivel de aula es de 0,7, todas aquellas aulas cuyo residuo de crecimiento este comprendido entre -0,07 y 0,07, son excluidas de los análisis de contingencia, ya que, se consideran que pueden tender a no diferenciarse de la media, no por su error de estimación, sino porque su valor en la pendiente hace que se sitúen ahí.

Con el tamaño muestral obtenido tras aplicar dicha restricción, se realiza una tabla de contingencia donde,se puede estudiar dentro de cada situación (*a, b y c*), acorde al tamaño de las aulas (pequeño, mediano y grande), el porcentaje de las mismas que se diferencian de la media y, por ende, las que no. Dichas tablas se acompañan de los estadísticos chi-cuadrado, Phi y V de Cramer para estudiar la relación entre el tamaño de aulas y su residuo de crecimiento categorizado como 0 = no se diferencia de la media y 1 = se diferencia de la media.

CAPÍTULO VI: PRESENTACIÓN DE RESULTADOS

En este capítulo, se presentan los resultados organizados acorde a cada uno de los dos problemas de investigación y, dentro de cada uno de ellos, se especifican los resultados relacionados con cada uno de los objetivos específicos definidos anteriormente (*ver apartado V.2*).

VI.1. Problema 1. Estudio de la fiabilidad en las estimaciones de las puntuaciones de los estudiantes y del comportamiento del ERM cuando se modifican elementos esenciales del modelo de crecimiento (varianza y correlación entre el intercepto y la pendiente)

Mediante el primer problema de investigación, se pretendía conocer el efecto de los porcentajes de varianza fijados en los coeficientes aleatorios, así como de los valores de correlación establecidos entre el intercepto y la pendiente, en la fiabilidad con la que se estiman las puntuaciones de los estudiantes a lo largo del tiempo, y en el comportamiento del efecto de regresión a la media.

Para dar respuesta a este problema de investigación, basado en una simulación Monte Carlo de carácter longitudinal y multinivel, se establecieron dos objetivos específicos. A través del primero de ellos, se pretendía garantizar una adecuada recuperación de las condiciones experimentales fijadas antes de la generación definitiva de los resultados. En el segundo objetivo se planteó estudiar el efecto de las condiciones experimentales manipuladas en los distintos modelos simulados, sobre la fiabilidad de las estimaciones de las puntuaciones de los estudiantes, y sobre el comportamiento del ERM, bajo tres situaciones distintas (*situación a, b y c*)

A continuación, se presentan los resultados obtenidos en cada uno de estos dos objetivos específicos.

VI.1.1. Objetivo específico 1. Garantizar una adecuada recuperación de las condiciones experimentales manipuladas

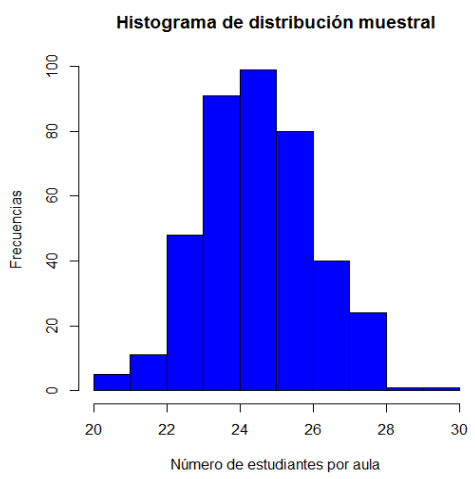
Con el fin de garantizar la fiabilidad y la validez de los resultados finales generados, en el apartado V.3.3.2 (*Precisión en la recuperación de parámetros*) del capítulo anterior, se describieron con detalle cada una de las pruebas y criterios establecidos para la consecución de dicho objetivo.

VI. 1.1.1. Recuperación del tamaño muestral

En cuanto a la recuperación del tamaño muestral fijado, tal y como se ha comentado en el apartado de descripción de variables (*apartado V.3.2*), no se fijó un número determinado de aulas, pero sí el número total de estudiantes (10.000), y la condición de que el mínimo de estudiantes por aula fueran 20 y el máximo 30, siendo la media de 25 estudiantes por aula.

Para el estudio de la precisión con la que se recuperan las condiciones muestrales fijadas previamente, en la siguiente tabla se especifica el número de alumnos por aula, el número de aulas acorde a un determinado número de estudiantes, así como el resultado de multiplicar, el número de estudiantes por el número de aulas. Todo ello acompañado de un histograma que representa la distribución muestral de todos los modelos simulados en el primer problema de investigación, ya que se fijó la misma semilla en todos ellos.

Tabla VI.1. *Recuperación de los tamaños muestrales*

Número de estudiantes por aula	Número de aulas	N estudiantes * N aulas	Gráfico
20	**1**	**20**	
21	4	84	
22	11	242	
23	48	1104	
24	91	2184	
25	**99**	**2475**	
26	80	2080	
27	40	1080	
28	24	672	
29	1	29	
30	**1**	**30**	
N Total	**400**	**10.000**	

Como se puede observar en la tabla anterior (Tabla VI.1), un aula está compuesta por 20 estudiantes, 4 aulas por 21 estudiantes, 11 aulas por 22 estudiantes, 48 aulas por 23 estudiantes, 91 aulas por 24 estudiantes, 99 aulas por 25 estudiantes, 80 aulas por 26 estudiantes, 40 aulas por 27 estudiantes, 24 aulas por 28 estudiantes, 1 aula por 29 estudiantes y 1 aula por 30 estudiantes., obteniéndose una muestra total de 10.000 sujetos repartidos en 400 aulas.

Acorde a estos resultados se puede considerar que existe una adecuada recuperación de las condiciones muestrales definidas previamente, ya que, se genera un total de 10.000 estudiantes repartidos en aulas formadas por un mínimo de 20 estudiantes y un máximo de 30, siendo las aulas compuestas por 25 estudiantes las que más presentes se encuentran en la distribución muestral.

VI.1.1.2. Recuperación de los parámetros de varianza y correlación

Para comprobar la precisión con la que se recuperan el resto de factores manipulados antes de la generación definitiva de los resultados, tal y como se explicó en la Tabla V.2, se calculó para cada condición experimental, el valor promedio obtenido en las mil réplicas, la desviación típica de la media, el sesgo, el coeficiente de variación y el boxplot de cada coeficiente generado en el promedio de las 1.000 réplicas.

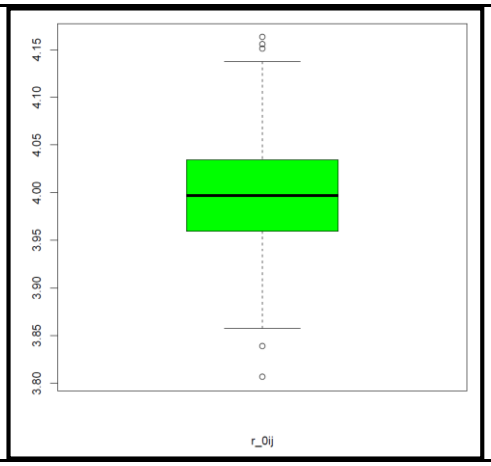
Debido a que se simularon casi 100 modelos diferentes (concretamente 99), en este apartado tan sólo se presenta la precisión con la que se recuperan los parámetros fijados en dos de los modelos simulados dentro de cada una de las situaciones (*situación a, b y c*), el resto de los resultados pueden ser consultados en el Anexo II.

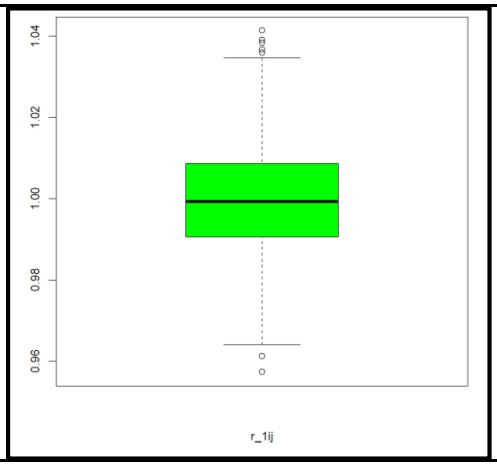
VI.1.1.2.1. Recuperación de los parámetros fijados en la situación a

Dentro de la ***situación a*** (se fija solo varianza en algunos coeficientes aleatorios), se muestran los resultados del modelo 7.3 y del modelo 7.4. En estos modelos la varianza del error intrasujeto se fijó a un 50% de la varianza total del modelo, la varianza del intercepto a un 40% y la de la pendiente a un 10%. Estos dos últimos coeficientes fueron fijados en el modelo 7.3 a nivel de estudiante y, en el modelo 7.4, a nivel de aula.

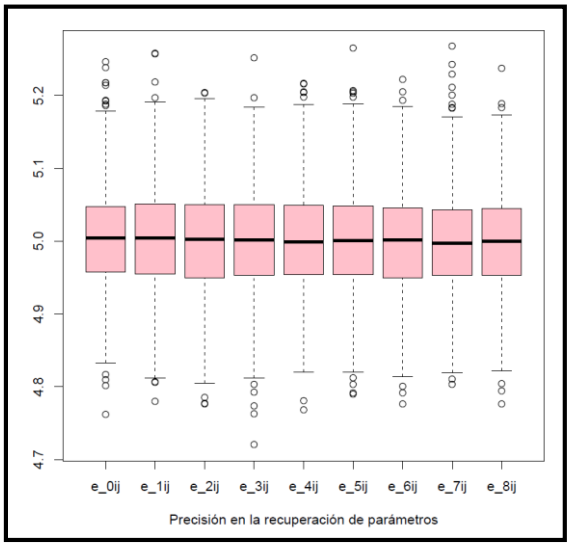
Tabla VI.2. *Precisión en la recuperación de parámetros del modelo 7.3 y 7.4*

MODELO 7.3									
NIVEL DE ESTUDIANTES									
$\sigma(r_{0ij})$	$\sigma(r_{0ij}\_gene)$	DT_ media	Sesgo	Coef_ variación	$\sigma(r_{1ij})$	$\sigma(r_{1ij}\_gene)$	DT_ media	Sesgo	Coef_ variación
4 (40%)	**3,997**	0,055	-0,003	**1,38%**	1 (10%)	**0,999**	0,014	-0,001	**1,39%**

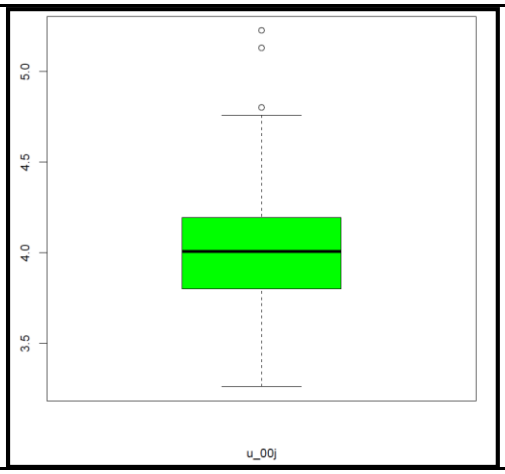


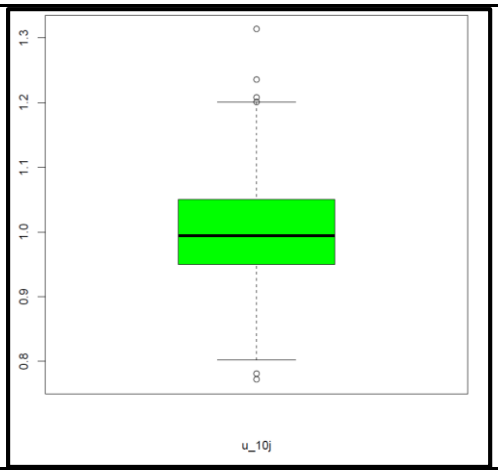


NIVEL DE TIEMPO											
$\sigma(e_{tij})$ = 5 (50%)											
$\sigma(e_{0ij})$	DT_ media	sesgo	CV	$\sigma(e_{1ij})$	DT_ media	sesgo	CV	$\sigma(e_{2ij})$	DT_ media	sesgo	CV
5,004	0,070	0,004	1,40%	**5,003**	0,071	0,003	1,42%	**5,000**	0,073	0,000	1,47%
$\sigma(e_{3ij})$	DT_ media	sesgo	CV	$\sigma(e_{4ij})$	DT_ media	sesgo	CV	$\sigma(e_{5ij})$	DT_ media	sesgo	CV
5,001	0,071	0,001	1,42%	**5,001**	0,071	0,001	1,41%	**5,001**	0,070	0,001	1,41%
$\sigma(e_{6ij})$	DT_ media	sesgo	CV	$\sigma(e_{7ij})$	DT_ media	sesgo	CV	$\sigma(e_{8ij})$	DT_ media	sesgo	CV
5,000	0,070	0,000	1,40%	**4,999**	0,069	-0,001	1,38%	**5,000**	0,067	0,000	1,33%

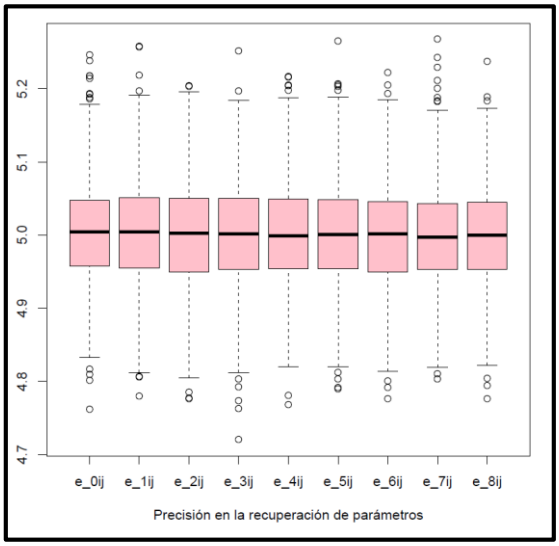


MODELO 7.4									
NIVEL DE AULAS									
$\sigma(u_{00j})$	$\sigma(u_{00j\_gene})$	**DT_ media**	**Sesgo**	**Coef_ variación**	$\sigma(u_{10j})$	$\sigma(u_{10j\_gene})$	**DT_ media**	**Sesgo**	**Coef_ variación**
4 (40%)	**4,011**	0,282	0,011	**7,04%**	1 (10%)	**0,999**	0,073	-0,001	**7,29%**





NIVEL DE TIEMPO											
$\sigma(e_{tij})$ = 5 (50%)											
$\sigma(e_{0ij})$	**DT_ media**	**sesgo**	**CV**	$\sigma(e_{1ij})$	**DT_ media**	**sesgo**	**CV**	$\sigma(e_{2ij})$	**DT_ media**	**sesgo**	CV
5,004	0,070	0,004	1,40%	**5,003**	0,071	0,003	1,42%	**5,000**	0,073	0,000	1,47%
$\sigma(e_{3ij})$	**DT_ media**	**sesgo**	**CV**	$\sigma(e_{4ij})$	**DT_ media**	**sesgo**	**CV**	$\sigma(e_{5ij})$	**DT_ media**	**sesgo**	**CV**
5,001	0,071	0,001	1,42%	**5,001**	0,071	0,001	1,41%	**5,001**	0,070	0,001	1,41%
$\sigma(e_{6ij})$	**DT_ media**	**sesgo**	**CV**	$\sigma(e_{7ij})$	**DT_ media**	**sesgo**	**CV**	$\sigma(e_{8ij})$	**DT_ media**	**sesgo**	**CV**
5,000	0,070	0,000	1,40%	**4,999**	0,069	-0,001	1,38%	**5,000**	0,067	0,000	1,33%



En el modelo 7.4, a nivel de estudiante se fijó un valor de varianza en el intercepto ($\sigma(r_{01j})$) de 4 (40% de la varianza total del modelo) y, se generó como promedio de las 1.000 réplicas ($\sigma\ (r_{01j\_gene})$) un valor de 3,999. En cuanto a la desviación típica de la media (DT_de la media) es muy pequeña de 0,055, por lo tanto, existe poca dispersión de los valores obtenidos en las mil réplicas respecto a la media, lo que indica que el estimador utilizado es eficiente, pues, la varianza del error es muy pequeña.

Debido a que existe muy poca diferencia entre el valor fijado (4) y el valor promedio obtenido (3,999), el sesgo también es muy pequeño, de -0,003, además al ser un valor negativo hay cierta infravaloración del parámetro. También se calculó el coeficiente de variación (CV) en términos de porcentaje, al obtenerse un valor tan pequeño como el 1,4%, indica una gran homogeneidad entre los valores de dicha variable. En cuanto al boxplot referido a dicho coeficiente, reafirma de manera más visual la precisión del parámetro generado, ya que la mediana es 4, el valor máximo es inferior a 4,15 y el valor mínimo superior a 3,85. Existen algunos valores atípicos en la parte superior del gráfico, que se encuentran en torno a 4,15 y en la parte inferior, el más alejado tiene un valor de 3,80.

Respecto a la precisión del valor fijado en la varianza de la pendiente ($\sigma(r_{10j})$), se fijó un valor de 1 (10% de la varianza total del modelo) y se generó un valor promedio de 0,999 ($\sigma\ (r_{10j}$_gene)), con una desviación promedia de 0,014, por lo tanto, dicho estimador es también muy eficiente. Como en la varianza del intercepto, el sesgo aquí también es negativo y muy cercano a 0. En cuanto al valor del coeficiente de variación indica que los valores obtenidos en las 1.000 réplicas de dicha variable son muy homogéneos (1,39%).

El valor fijado en la varianza en el error intrasujeto ($\sigma\ (e_{tij})$) se fijó a 5 (50% de la varianza total), y se generaron nueve valores de varianza, cada uno de ellos referido a una ocasión de medida distinta, pero en todos ellos el valor promedio obtenido es muy cercano a 5, por ello el sesgo es prácticamente 0. En cuanto al coeficiente de variación, en las diferentes ocasiones de medida se encuentra entre el 1,40% y 1,47%, por lo tanto, los valores de varianza del error intrasujeto, en cada una de las aplicaciones, a lo largo de las mil réplicas, son muy similares.

Si se comparan estos resultados con los generados para el modelo 7.4, se observan algunas diferencias. Aunque en el modelo 7.4 el valor de la varianza del intercepto generada en las 1.000 réplicas (4,011) es similar al fijado (4), la desviación típica es muy superior (0,282) a la del modelo 7.3 (0,055), lo que indica una mayor dispersión entre los valores de dicha variable y, por lo tanto, una menor eficiencia. Esto conlleva a que el coeficiente de variación de dicho modelo 7.4 (7,04%), sea mayor al del modelo 7.3 (1,38%). Algo similar ocurre al comparar la varianza de la pendiente generada en el nivel de aula (modelo 7.4), con la varianza de la pendiente generada en el nivel de estudiante (modelo 7.3), aunque en ambos modelos, los promedios son los mismos (0,999), la desviación típica de la media es mayor en el modelo 7.4 y, por ello, el coeficiente de variación también lo es.

Que exista una mayor dispersión en el modelo 7.4, en el cual se fijó variabilidad en el intercepto y la pendiente del nivel de aula, respecto al modelo 7.3 donde se estableció valores de varianza en los mismos coeficientes, pero referidos al nivel de estudiante, se puede deber a la diferencia en el número de unidades de cada nivel. Pues mientras que en el modelo 7.4 se generan los parámetros para 400 aulas, en el modelo 7.3 se trabajó con10.000 estudiantes.

Finalmente, en cuanto a las varianzas del error intrasujeto en ambos modelos se obtienen los mismos resultados, ya que se fija los mismos porcentajes de varianza en el mismo nivel.

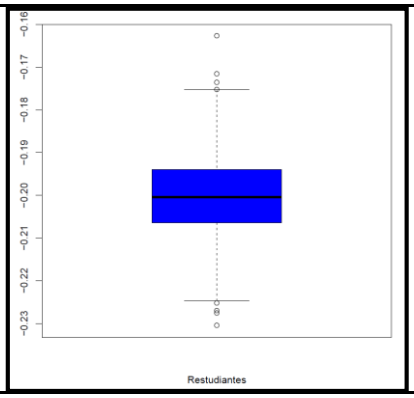
VI.1.1.2.2. Recuperación de los parámetros fijados en la situación b

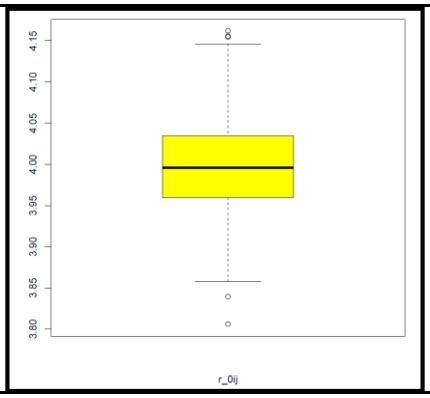
Dentro de la ***situación b*****,** en la cual se fija varianza en el error intrasujeto, en el intercepto y en la pendiente de un mismo nivel (aula o estudiante) y, distintos valores de correlación entre el intercepto y la pendiente de uno de esos niveles, se presentan los resultados obtenidos en el modelo 11.3 y en el modelo 11.4.

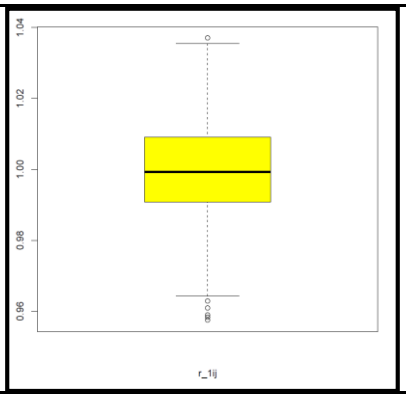
En estos modelos la varianza del error intrasujeto se fijó a un 50%, la varianza del intercepto a un 40%, y la varianza de la pendiente a un 10%. En el modelo 11.3, estos dos últimos coeficientes aleatorios se fijaban en el nivel de estudiante, además en dicho nivel se fijó una correlación entre el intercepto y la pendiente de -0,2. En el modelo 11.4 se fijó también dicho valor de correlación, pero a nivel de aula, ya que, se estableció variabilidad en el estatus (40%) y en la pendiente (10%) de dicho nivel.

Tabla VI.3. *Precisión en la recuperación de parámetros del modelo 11.3 y del modelo 11.4*

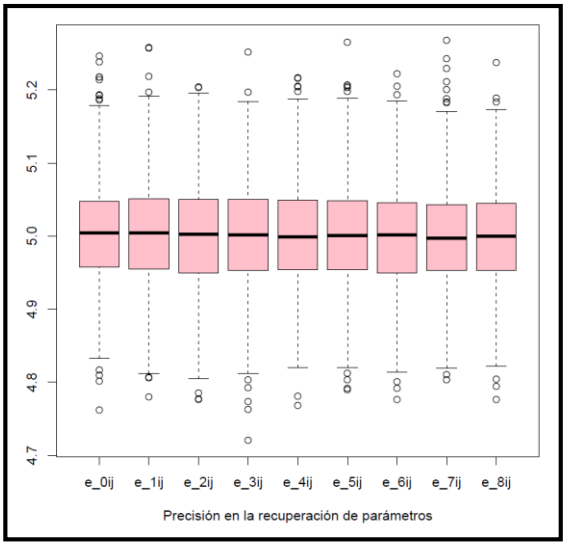
MODELO 11.3									
NIVEL DE ESTUDIANTES									
$R_{estudiantes}$		$R_{estudiantes_}$gene			DT_réplicas		Sesgo		Coef_variación
-0,2		-0,200			0,010		0,000		4,86%
$\sigma(r_{0ij})$	$\sigma(r_{0ij\_}gene)$	DT_ media	Sesgo	Coef_ variación	$\sigma(r_{1ij})$	$\sigma(r_{1ij\_}gene)$	DT_ media	Sesgo	Coef_ variación
4 (40%)	**4,002**	0,056	0,002	1,41%	1 (10%)	**1,000**	0,014	0,000	1,40%







NIVEL DE TIEMPO											
$\sigma(e_{tij})$ = 5 (50%)											
$\sigma(e_{0ij})$	DT_ media	sesgo	CV	$\sigma(e_{1ij})$	DT_ media	sesgo	CV	$\sigma(e_{2ij})$	DT_ media	sesgo	CV
5,001	0,070	0,001	1,40%	**5,006**	0,072	0,006	1,43%	**4,995**	0,073	-0,005	1,46%
$\sigma(e_{3ij})$	DT_ media	sesgo	CV	$\sigma(e_{4ij})$	DT_ media	sesgo	CV	$\sigma(e_{5ij})$	DT_ media	sesgo	CV
5,001	0,072	0,001	1,45%	**4,998**	0,070	-0,002	1,40%	**5,005**	0,070	0,005	1,41%
$\sigma(e_{6ij})$	DT_ media	sesgo	CV	$\sigma(e_{7ij})$	DT_ media	sesgo	CV	$\sigma(e_{8ij})$	DT_ media	sesgo	CV
5,000	0,071	0,000	1,42%	**5,001**	0,069	0,001	1,39%	**4,999**	0,069	-0,001	1,38%

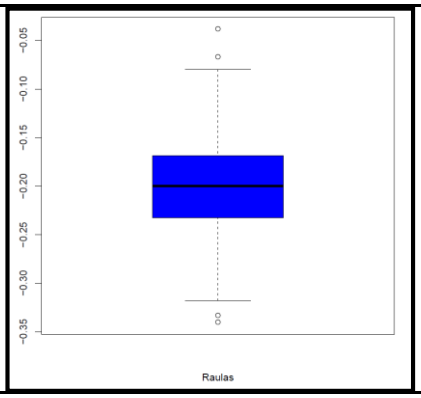


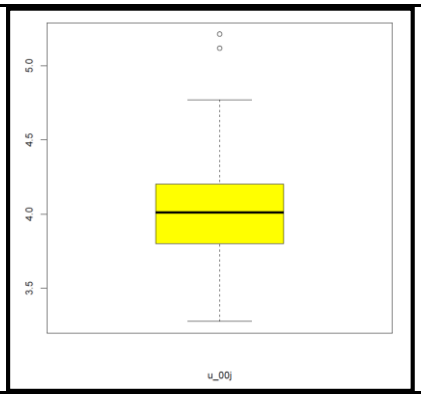
MODELO 11.4

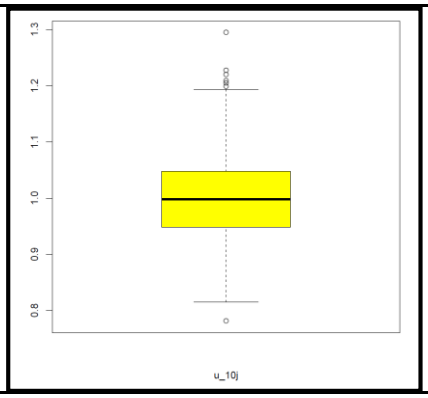
NIVEL DE AULA

R_{aulas}	R_{aulas}_gene	DT-media	Sesgo	Coef_variación
-0,2	-0,204	0,047	-0,004	23,18%

$\sigma(u_{00j})$	$\sigma(u_{00j}\_gene)$	DT_ media	Sesgo	Coef_ variación	$\sigma(u_{10j})$	$\sigma(u_{10j}\_gene)$	DT_ media	Sesgo	Coef_ variación
4 (40%)	3,989	0,289	-0,011	7,24%	1 (10%)	0,997	0,072	-0,003	7,243%



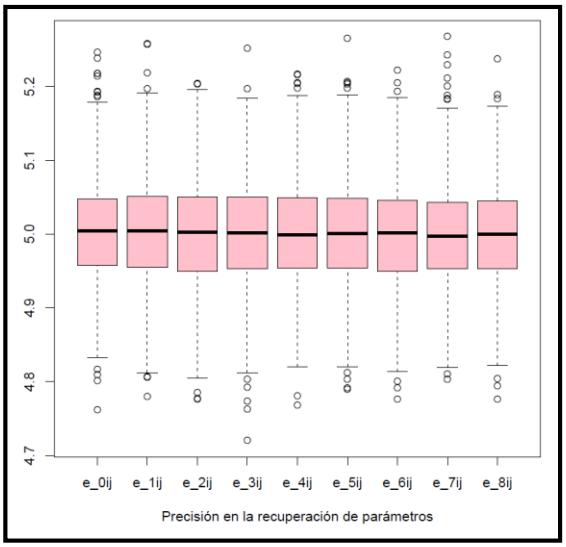




NIVEL DE TIEMPO

$\sigma(e_{tij})$ = 5 (50%)

$\sigma(e_{0ij})$	DT_ media	sesgo	CV	$\sigma(e_{1ij})$	DT_ media	sesgo	CV	$\sigma(e_{2ij})$	DT_ media	sesgo	CV
5,001	0,070	0,001	1,40%	5,006	0,072	0,006	1,43%	4,995	0,073	-0,005	1,46%
$\sigma(e_{3ij})$	DT_ media	sesgo	CV	$\sigma(e_{4ij})$	DT_ media	sesgo	CV	$\sigma(e_{5ij})$	DT_ media	sesgo	CV
5,001	0,072	0,001	1,45%	4,998	0,070	-0,002	1,40%	5,005	0,070	0,005	1,41%
$\sigma(e_{6ij})$	DT_ media	sesgo	CV	$\sigma(e_{7ij})$	DT_ media	sesgo	CV	$\sigma(e_{8ij})$	DT_ media	sesgo	CV
5,000	0,071	0,000	1,42%	5,001	0,069	0,001	1,39%	4,999	0,069	-0,001	1,38%



En el modelo 11.3 la correlación entre el intercepto y la pendiente del nivel de estudiante se fijó a -0,2 y, se genera un valor promedio en las 1.000 réplicas de -0,2 con una desviación típica de 0,010. En el modelo 11.4 también se fijó ese mismo valor de correlación, pero referido al nivel de aula. En dicho modelo, el valor medio obtenido es también de -0,2, pero la desviación típica de la media es de 0,047, superior a la del modelo 11.3, lo que indica una mayor dispersión en los valores de esta variable en el modelo 11.4 y, por lo tanto, una menor eficiencia de dicho parámetro. Como consecuencia de esta mayor variabilidad, el coeficiente de variación de la correlación entre el intercepto y la pendiente, en el modelo 11.4 es casi cuatro veces más (23,18%), al coeficiente de variación obtenido en el modelo 11.3 (4,86%).

Esta misma tendencia ocurre cuando se compara los valores generados en la varianza del intercepto del modelo 11.3 con los del modelo 11.4, o cuando se compara los resultados de los valores generados de la varianza de la pendiente entre ambos modelos.

En el modelo 11.4, existe una mayor dispersión de los valores obtenidos respecto a la media, que en el modelo 11.3. Esto no quiere decir que en el modelo 11.4 exista una mala recuperación de los parámetros fijados, sino que la recuperación es menos precisa que la del modelo 11.3, donde se trabaja con un mayor número de unidades. Además, si en ambos modelos se comparan los boxplot referidos al coeficiente de correlación entre el intercepto y la pendiente, el coeficiente de la varianza del intercepto y de la varianza de la pendiente, las medianas son muy similares, aunque los valores extremos son superiores en el modelo 11.4.

La existencia de una menor dispersión en los parámetros generados en el nivel de estudiante, respecto a los del nivel de aula, se puede justificar nuevamente debido a la diferencia en el número de unidades muestrales.

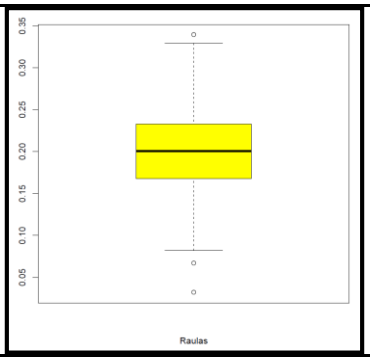
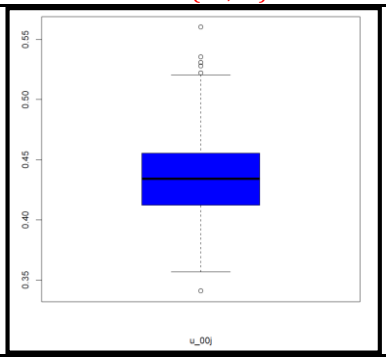
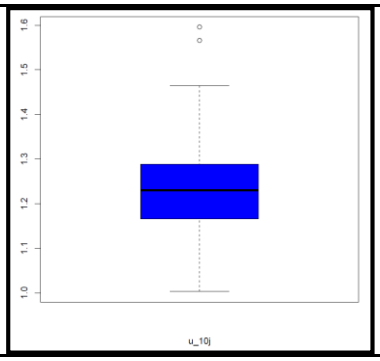
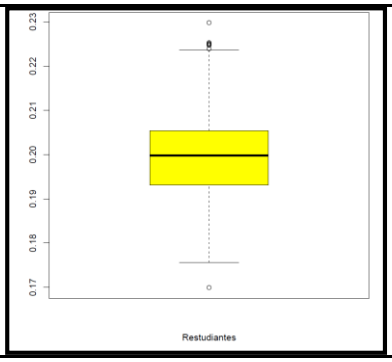
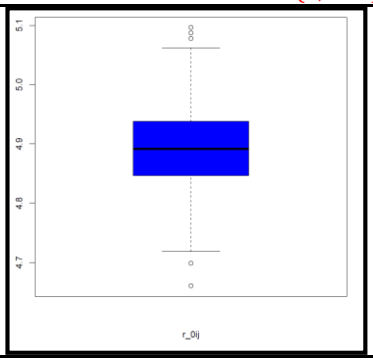
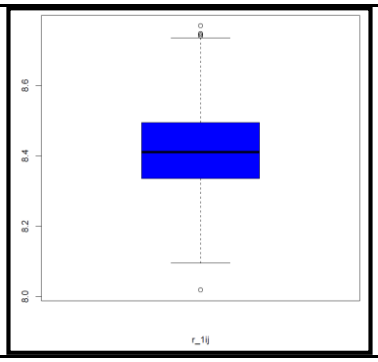
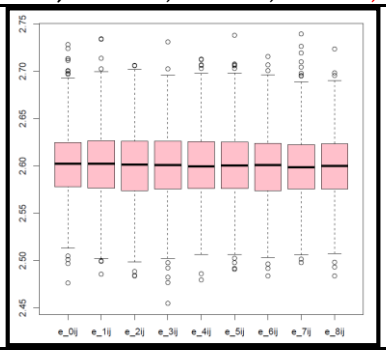
VI.1.1.2.2. Recuperación de los parámetros fijados en la situación c

Dentro de la ***situación c***, caracterizada porque se fijó varianza en todos los coeficientes aleatorios y un valor de correlación entre el intercepto y la pendiente, tanto del nivel de estudiante, como del nivel de aula. Se toman como ejemplo los resultados del modelo 17, referidos a la combinación 1 y combinación 6.

Recuérdese que en dicho modelo 17 la varianza del intercepto del nivel de estudiante se fijó a un 48,96%, la varianza de la pendiente de ese mismo nivel a un 8,42%, la varianza del estatus del nivel de aula a un 4,35% y la varianza de la pendiente a un 12,27%. El 26% de varianza restante se fijó en la varianza del error intrasujeto.

Dentro de este modelo, en la combinación 1, tanto la correlación entre el intercepto y la pendiente del nivel de estudiante, como la del nivel de aula, se fijó a 0,2. Sin embargo, en la combinación 6 ambas correlaciones se fijaron a 0,9.

Tabla VI.4. *Precisión en la recuperación de parámetros del modelo 17 combinación 1 y 6*

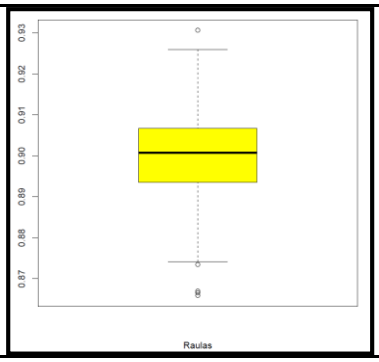
MODELO 17 COMBINACIÓN 1										
NIVEL DE AULA										
R_{aulas}		R_{aulas}_gene		DT-media		Sesgo			Coef_variación	
0,20		0,200		0,047		0,000			23,26%	
$\sigma(u_{00j})$	$\sigma(u_{00j}\_gene)$	DT_ media	Sesgo	Coef_ variación	$\sigma(u_{10j})$	$\sigma(u_{10j}\_gene)$	DT_ media	Sesgo	Coef_ variación	
0,435 (4,35%)	0,434	-0,001	0,031	7,18%	1,227 (12,27)	1,223	0,086	-0,004	7,05%	
										
NIVEL DE ESTUDIANTES										
$R_{estudiantes}$		$R_{estudiantes}$_gene		DT_réplicas		Sesgo			Coef_variación	
0,2		0,200		0,010		0,000			5,03%	
$\sigma(r_{0ij})$	$\sigma(r_{0ij}\_gene)$	DT_ media	Sesgo	Coef_ variación	$\sigma(r_{1ij})$	$\sigma(r_{1ij}\_gene)$	DT_ media	Sesgo		
4,896 (48,96%)	4,895	0,070	-0,001	1,44%	0,842 (8,42%)	0,842	0,012	0,000		
										
NIVEL DE TIEMPO										
$\sigma(e_{tij})$ = 2,6 (26%)										
$\sigma(e_{0ij})$	DT_ media	sesgo	CV	$\sigma(e_{1ij})$	DT_ media	sesgo	CV	$\sigma(e_{2ij})$	DT_ media	sesgo	CV
2,599	0,038	-0,001	1,46%	2,599	0,037	-0,001	1,41%	2,599	0,037	-0,001	1,42%
$\sigma(e_{3ij})$	DT_ media	sesgo	CV	$\sigma(e_{4ij})$	DT_ media	sesgo	CV	$\sigma(e_{5ij})$	DT_ media	sesgo	CV
2,603	0,036	0,003	1,39%	2,599	0,036	-0,001	1,40%	2,599	0,038	-0,001	1,47%
$\sigma(e_{6ij})$	DT_ media	sesgo	CV	$\sigma(e_{7ij})$	DT_ media	sesgo	CV	$\sigma(e_{8ij})$	DT_ media	sesgo	CV
2,599	0,037	-0,001	1,42%	2,600	0,037	0,000	1,40%	2,601	0,037	0,001	1,41%
										

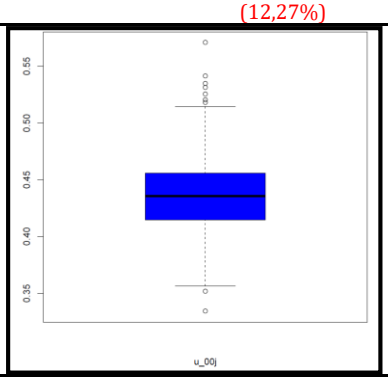
MODELO 16 COMBINACIÓN 6

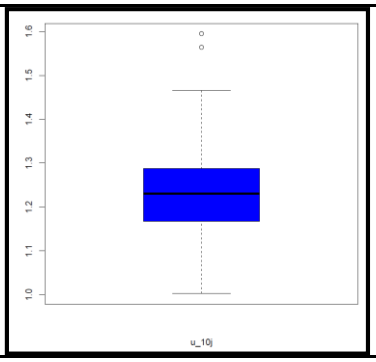
NIVEL DE AULA

R_{aulas}	R_{aulas}-gene	DT-media	Sesgo	Coef_variación
0,9	0,900	0,009	0,000	1,03%

$\sigma(u_{00j})$	$\sigma(u_{00j}\_gene)$	DT_ media	Sesgo	Coef_ variación	$\sigma(u_{10j})$	$\sigma(u_{10j}\_gene)$	DT_ media	Sesgo	Coef_ variación
0,435 (4,35%)	0,434	0,030	-0,001	6,99%	1,227 (12,27%)	1,226	0,085	-0,001	6,92%



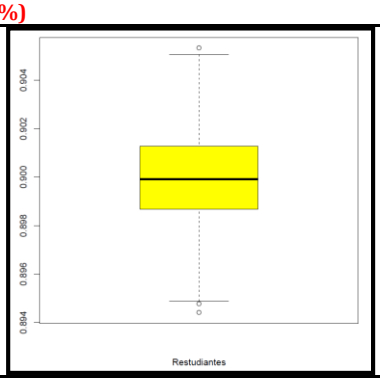


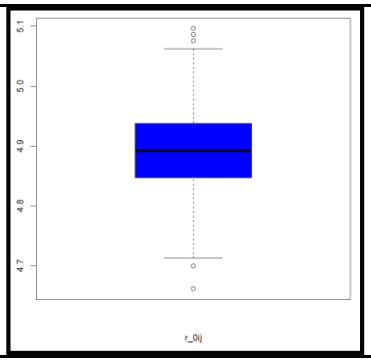


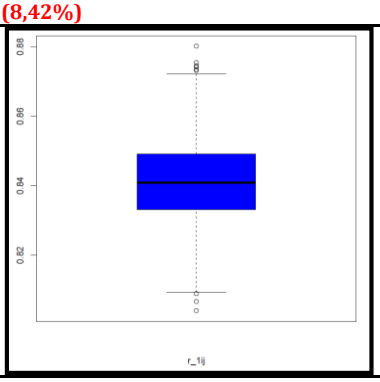
NIVEL DE ESTUDIANTES

$R_{estudiantes}$	$R_{estudiantes}$-gene	DT_réplicas	Sesgo	Coef_variación
0,9	0,900	0,002	0,000	0,21%

$\sigma(r_{0ij})$	$\sigma(r_{0ij}\_gene)$	DT_ media	Sesgo	Coef_ variación	$\sigma(r_{1ij})$	$\sigma(r_{1ij}\_gene)$	DT_ media	Sesgo
4,896 (48,96%)	4,899	0,068	0,003	1,39%	0,842 (8,42%)	0,842	0,012	0,000







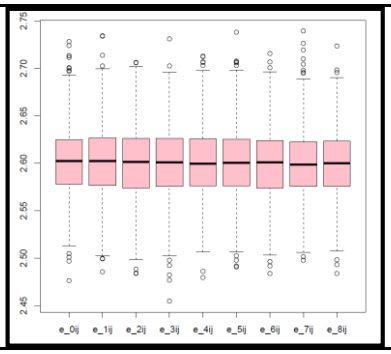
NIVEL DE TIEMPO

$\sigma(e_{tij})$ = 2,6(26%)

$\sigma(e_{0ij})$	DT_ media	sesgo	CV	$\sigma(e_{1ij})$	DT_ media	sesgo	CV	$\sigma(e_{2ij})$	DT_ media	sesgo	CV
2,598	0,036	-0,002	1,40%	2,602	0,037	0,002	1,40%	2,601	0,038	0,001	1,45 %

$\sigma(e_{3ij})$	DT_ media	sesgo	CV	$\sigma(e_{4ij})$	DT_ media	sesgo	CV	$\sigma(e_{5ij})$	DT_ media	sesgo	CV
2,601	0,039	0,001	1,48%	2,599	0,039	-0,001	1,48%	2,601	0,038	0,001	1,47 %

$\sigma(e_{6ij})$	DT_ media	sesgo	CV	$\sigma(e_{7ij})$	DT_ media	sesgo	CV	$\sigma(e_{8ij})$	DT_ media	sesgo	CV
2,600	0,037	0,000	1,42%	2,602	0,036	0,002	1,38%	2,600	0,037	0,000	1,43 %



Como se puede observar, nuevamente, si se compara dentro de un mismo modelo los valores generados en la correlación entre el intercepto y la pendiente del nivel de aula, con el del nivel de estudiante, existe menos dispersión en el nivel de estudiante.

Además, se observa que cuando el valor fijado en la correlación entre el intercepto y la pendiente es bajo (modelo 17, combinación 1), existe una mayor dispersión en el coeficiente de variación que cuando se fija un valor de correlación entre el intercepto y la pendiente alta (modelo 17, combinación 6)

Del mismo modo, si se comparan los coeficientes aleatorios del intercepto y la pendiente entre ambos modelos, a pesar de que se fija el mismo porcentaje de varianza y la media es similar, se observa que la dispersión es ligeramente superior en el modelo de la combinación 1, sobre todo en el nivel de aula, respecto al modelo 17 de la combinación 6. Esto puede ser debido al efecto de la correlación fijada entre el intercepto y la pendiente, puesto que la media es inferior en la combinación 1.

A la vista de todo esto, se puede concluir que tanto el programa utilizado como la programación realizada es adecuada para la generación de los resultados finales, ya que, tal y como se ha demostrado y, acorde con los valores obtenidos en los modelos presentados en el Anexo II, existe una adecuada recuperación de los factores manipulados, constatado por la realización de diversas pruebas estadísticas. Además, las mismas revelaron que los parámetros generados a nivel de estudiante son más precisos que los simulados en el nivel de aula, debido a que este último nivel cuenta con un tamaño muestral inferior, asimismo el valor de correlación fijado entre el intercepto y la pendiente también afecta en la dispersión de las distintas variables, siendo inferior esta dispersión cuanto mayor es la intensidad de relación.

VI.1.2. Objetivo específico 2. Estudio de la fiabilidad de las puntuaciones de los estudiantes y del comportamiento del ERM bajo distintas condiciones experimentales

Con el fin de conocer la fiabilidad en las estimaciones de las puntuaciones de los estudiantes en evaluaciones longitudinales, así como el comportamiento del ERM; a continuación, se presentan las matrices de correlaciones generadas en los distintos modelos simulados.

Dichas matrices se encuentran organizadas de acuerdo a las distintas situaciones (*situación a, b y c*), que se han explicado en el apartado *V.3.3.3. (Procedimiento de generación de modelos).* En las mismas se analiza tanto el patrón existente entre la primera ocasión de medida y el resto de ocasiones, que hace referencia al comportamiento del efecto de regresión a la media, como el patrón obtenido entre el resto de mediciones.

Cómo se ha comentado en la metodología, no se presentan los resultados de cada una de las 1.000 réplicas realizadas para cada modelo simulado, sino, las correlaciones promedias de esas 1.000 réplicas en cada uno de ellos. En estas matrices de correlación su intensidad se interpreta siguiendo la clasificación de Wonnacott y Wonnacott (1999), donde una correlación menor a 0,20 se considera muy baja, entre 0,20 y 0,40 baja, entre 0,40 y 0,70 moderada, entre 0,70 y 0,90 alta y mayor a 0,90 muy alta.

Los resultados se encuentran organizados en tablas, donde en la parte superior se especifica el modelo al que se refiere. En la parte izquierda aparecen los coeficientes fijos (β), donde β_{000} y β_{001} son constantes, siempre toman los mismos valores (4,6 y 0,4 respectivamente), y el porcentaje fijado en los diferentes coeficientes aleatorios, organizados por niveles. El nivel tres hace referencia a las aulas, en el mismo se especifica el porcentaje fijado en la varianza del estatus $\left(\sigma\ (u_{00j})\right)$ y en la varianza de la pendiente $\left(\sigma(u_{1ij})\right)$, y, los mismos coeficientes para el nivel dos, referido al estudiante $\left(\sigma(r_{0ij});\ \sigma(r_{1ij})\right)$. Finalmente, en el nivel uno (tiempo) se especifica la varianza del error intrasujeto $(\sigma(e_{tij}))$.

En la parte derecha de la tabla aparece la matriz de correlación de las puntuaciones de los estudiantes en las distintas ocasiones de medida. Se trata de una

matriz de 9x9, donde tanto en la cabecera como en el lateral izquierdo se especifica de la siguiente manera: $Y_{0ij,}$ $Y_{1ij}, Y_{2ij}, Y_{3ij}, Y_{4ij}, Y_{5ij}, Y_{6ij}, Y_{7ij}, Y_{8ij}$ la ocasión de medida correspondiente, siendo $Y_{0ij,}$ la medida de la ocasión 1, Y_{1ij}, la ocasión 2 y así sucesivamente hasta la medida nueve (Y_{8ij})

También, junto con el valor de correlación promedio obtenido en las 1.000 réplicas, de las diferentes ocasiones de medida y modelos, se estimó la desviación típica promedia, pero en todos los casos es 0 o muy cercana a dicho valor, por ello, se ha decidido omitir dicho dato.

Asimismo, aquellos resultados que se consideran más relevantes por el patrón de correlación generado, se acompañan de un gráfico que muestra dicho patrón de manera más visual. En estos gráficos, donde el valor de correlación aparece representado por un círculo, es importante a la hora de su interpretación tener en cuenta, tanto su tamaño, como su color. El color hace referencia a la direccionalidad de la correlación y a la intensidad de la misma. Un color rojo o de tonalidad similar, indica una correlación negativa, mientras que el azul o un color afín, una correlación positiva. A medida que se intensifican dichos colores es que la correlación tiende a ser perfecta, en el caso de que el círculo sea de color rojo será perfecta negativa, y perfecta positiva si es azul.

El tamaño de los círculos también es otro indicador de la fuerza de asociación existente entre dos mediciones. Cuanto más grande sea el círculo mayor es la intensidad entre las mediciones referidas, por el contrario, cuanto más pequeño sea el círculo menos relación existe entre las aplicaciones, pudiendo llegar a ser nula, ya que el gráfico representa todos los valores de correlación comprendidos entre 1 y -1.

VI.1.2.1. Resultados de los modelos de la situación a

En este *apartado VI.1.2.1.* se presentan las matrices de correlaciones obtenidas en los modelos desarrollados dentro de la ***situación a***, es decir, cuando solo se fija variabilidad en algunos de los coeficientes aleatorios del modelo.

Dentro de esta situación, se desarrolló un primer modelo (**modelo 1**), en el cual, recuérdese, sólo se fijó variabilidad en el error-intrasujeto, por lo que se asumía que no había efecto ni de las características de las aulas, ni de los estudiantes sobre los niveles de logro.

En este modelo el patrón de correlación obtenido es el siguiente:

Tabla VI.5. *Matriz de correlaciones del modelo 1*

MODELO 1

EFECTOS FIJOS		
Coef.		
β_{000}	**4,6**	
β_{100}	**0,4**	
EFECTOS ALEATORIOS		
NIVEL 3	$\sigma(u_{00j})$	$\sigma(u_{10j})$
$\sigma(u_{00j})$	0	
$\sigma(u_{10j})$		0
NIVEL 2	$\sigma(r_{0ij})$	$\sigma(r_{1ij})$
$\sigma(r_{0ij})$	0	
$\sigma(r_{1ij})$		0
NIVEL 1	$\sigma(e_{tij})$	
$\sigma(e_{tij})$	**100%**	
R_{aulas}	0	
$R_{estudiantes}$	0	

MATRIZ DE CORRELACIONES

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	-0,0194	-0,0035	0,0089	0,0059	0,0067	0,0099	-0,0076	0,0002
Y_{1ij}	-0,0194	1	0,0021	-0,005	0,0128	-0,0114	0,0039	-0,0042	0,0066
Y_{2ij}	-0,0035	0,0021	1	0,0002	-0,0137	-0,006	-0,0028	-0,01	0,0107
Y_{3ij}	0,0089	-0,005	0,0002	1	0,0009	-0,002	0,0142	-5,00E-04	0,0204
Y_{4ij}	0,0059	0,0128	-0,0137	0,0009	1	0,0014	0,0097	-0,0028	-0,004
Y_{5ij}	0,0067	-0,0114	-0,006	-0,002	0,0014	1	0,0039	-0,0164	0,0008
Y_{6ij}	0,0099	0,0039	-0,0028	0,0142	0,0097	0,0039	1	0,0117	0,009
Y_{7ij}	-0,0076	-0,0042	-0,01	-5,00E-04	-0,0028	-0,0164	0,0117	1	**-0,0231**
Y_{8ij}	0,0002	0,0066	0,0107	0,0204	-0,004	0,0008	0,009	-0,0231	1

Como se puede observar en la tabla anterior VI.5, la correlación entre la primera ocasión de medida y el resto de mediciones es constante, en torno a 0. Por lo tanto, siguiendo a Nesselroade et al. (1980) el efecto de regresión a la media que se produce entre la primera ocasión con el resto de ocasiones es constante, no varía según se incrementa el número de mediciones.

Este patrón también se mantiene constante (en torno a 0) en el resto de correlaciones, incluso cuando el número de mediciones es elevado. Por lo tanto, cuando toda la variabilidad se fija en el error intrasujeto, incluso ni un número muy elevado de mediciones, como son nueve ocasiones (tal y como se muestra en negrita), proporcionan información fiable sobre la puntuación verdadera del rendimiento de los estudiantes, ya que, los valores de correlación obtenidos entre las distintas ocasiones de medida son nulos.

En las siguientes tablas (VI.6, VI.7, VI.8) se presentan los **modelos de la subsituación 2**. En la cual se fijaban diferentes porcentajes de varianza en el error intrasujeto y en la pendiente del nivel de estudiante, pero no en el nivel de aula, lo que supone que el rendimiento de los estudiantes a lo largo del tiempo se debe a las características sistemáticas de los mismos y no a la de las aulas.

A continuación, se muestra la matriz de correlaciones obtenida en las puntuaciones de los estudiantes entre las distintas mediciones en el **modelo 2**.

Tabla VI.6. *Matriz de correlaciones del modelo 2*

MODELO 2		
EFECTOS FIJOS		
Coef.		
β_{000}	**4,6**	
β_{100}	**0,4**	
EFECTOS ALEATORIOS		
NIVEL 3	$\sigma(u_{00j})$	$\sigma(u_{10j})$
$\sigma(u_{00j})$	0	
$\sigma(u_{10j})$		0
NIVEL 2	$\sigma(r_{0ij})$	$\sigma(r_{1ij})$
$\sigma(r_{0ij})$	0	
$\sigma(r_{1ij})$		**5%**
NIVEL 1	$\sigma(e_{tij})$	
$\sigma(e_{tij})$	**95%**	
R_{aulas}	0	
$R_{estudiantes}$	0	

MATRIZ DE CORRELACIONES

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**-0,0208**	**-0,0066**	**0,0027**	**-0,0013**	**-0,0017**	**-0,0008**	**-0,011**	**-0,0071**
Y_{1ij}	-0,0208	1	**0,0944**	0,1225	0,1615	0,1616	0,1849	0,19	0,2022
Y_{2ij}	-0,0066	0,0944	1	0,228	0,2649	0,3009	0,3277	0,3414	0,3635
Y_{3ij}	0,0027	0,1225	0,228	1	0,3754	0,4153	0,457	0,4731	0,4986
Y_{4ij}	-0,0013	0,1615	0,2649	0,3754	1	0,5019	0,546	0,5686	0,5891
Y_{5ij}	-0,0017	0,1616	0,3009	0,4153	0,5019	1	0,6031	0,6269	0,6559
Y_{6ij}	-0,0008	0,1849	0,3277	0,457	0,546	0,6031	1	0,6874	0,7112
Y_{7ij}	-0,011	0,19	0,3414	0,4731	0,5686	0,6269	0,6874	1	**0,7394**
Y_{8ij}	-0,0071	0,2022	0,3635	0,4986	0,5891	0,6559	0,7112	0,7394	1

Cuando en general todos los estudiantes parten del mismo nivel de rendimiento (no se fija varianza en el intercepto del nivel de estudiante) y todas las aulas tienen el mismo nivel de logro previo (no se fija varianza en el estatus del nivel de aula) se observa que el ERM se mantiene constante a lo largo del tiempo, tal y como reflejan los valores de la primera ocasión de medida que están en negrita.

En cuanto al patrón obtenido en el resto de correlaciones, aunque el porcentaje fijado en la varianza de la pendiente (5%), es muy pequeño en comparación con el porcentaje fijado en la varianza del error intrasujeto (95%), a partir de la ocasión dos y tres, los valores de correlación crecen dando lugar a un patrón acumulativo. Donde a medida que se incrementan el número de aplicaciones, las correlaciones entre las mismas son cada vez más elevadas y, por lo tanto, las puntuaciones generadas en dichas ocasiones son más fiables.

Este patrón acumulativo se debe, a que, si se recuerda, en la ecuación de regresión explicada en el apartado III.4.1. sobre *Modelos Jerárquicos Lineales completamente anidados*, los coeficientes referidos a los residuos vinculados a la pendiente del nivel de estudiante (r_{1ij}) y a la pendiente del nivel de aula (u_{10j}) que expresan las diferencias de cada estudiante y de cada aula, respectivamente, en relación a la media global, se encuentran multiplicados por la ocasión de medida correspondiente menos 1, como la varianza de la pendiente del nivel de estudiante y del tiempo.

Por lo tanto, cuando se multiplica dicho valor de la varianza de la pendiente por ocho, que corresponde a la ocasión nueve, aunque el valor de la varianza sea muy pequeño, como

en este caso de un 5%, respecto a la varianza total del modelo. El resultado es la obtención de un valor de correlación correspondiente a la ocasión ocho y nueve de 0,739, muy elevado en comparación con la correlación entre la ocasión dos y tres (0,094) (valores señalados en la tabla con negrita).

Así, en este modelo, aunque con tres ocasiones de medida no se estiman las puntuaciones de los estudiantes de manera muy fiable (la correlación es baja, en torno a 0,20), con nueve ocasiones la fiabilidad de las estimaciones se incrementa notablemente.

En la siguiente tabla (VI. 7) se presentan los resultados del **modelo 2.1,** en el cual se ha incrementado el porcentaje de varianza fijado en la pendiente del nivel de estudiante respecto al modelo anterior.

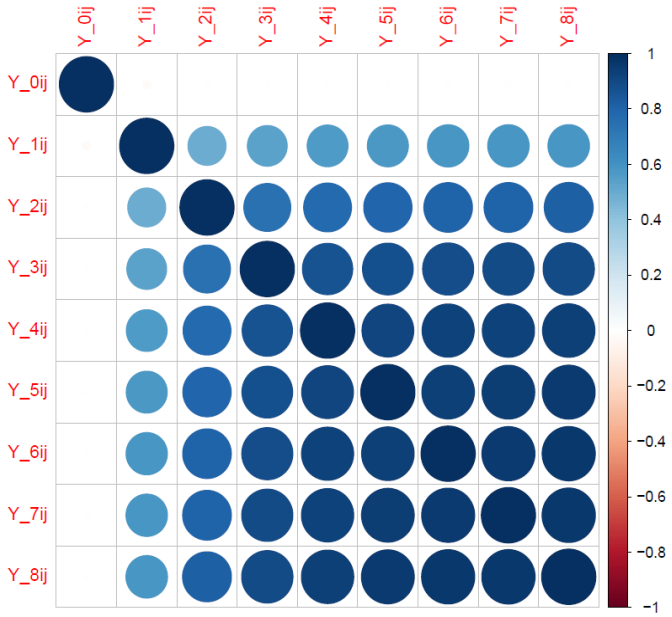
Tabla VI.7. *Matriz de correlaciones del modelo 2.1*

MODELO 2.1

EFECTOS FIJOS										
Coef.										
β_{000}	**4,6**									
β_{100}	**0,4**									
EFECTOS ALEATORIOS										
NIVEL 3	$\sigma(u_{00j})$	$\sigma(u_{10j})$								
$\sigma(u_{00j})$	0									
$\sigma(u_{10j})$		0								
NIVEL 2	$\sigma(r_{0ij})$	$\sigma(r_{1ij})$								
$\sigma(r_{0ij})$	0									
$\sigma(r_{1ij})$		**35%**								
NIVEL 1	e_{tij}									
$\sigma(e_{tij})$	**65%**									
R_{aulas}	0									
$R_{estudiantes}$	0									

MATRIZ DE CORRELACIONES

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**-0,0205**	**-0,0088**	**-0,0038**	**-0,0059**	**-0,0062**	**-0,0058**	**-0,0095**	**-0,0081**
Y_{1ij}	-0,0205	1	**0,4912**	0,5387	0,5663	0,5707	0,5805	0,5831	0,5872
Y_{2ij}	-0,0088	0,4912	1	0,7489	0,7777	0,7934	0,8031	0,8079	0,8135
Y_{3ij}	-0,0038	0,5387	0,7489	1	0,8602	0,8754	0,8871	0,8917	0,8971
Y_{4ij}	-0,0059	0,5663	0,7777	0,8602	1	**0,9123**	0,9235	0,9289	0,933
Y_{5ij}	-0,0062	0,5707	0,7934	0,8754	0,9123	1	0,9397	0,945	0,9501
Y_{6ij}	-0,0058	0,5805	0,8031	0,8871	0,9235	0,9397	1	0,9573	0,9613
Y_{7ij}	-0,0095	0,5831	0,8079	0,8917	0,9289	0,945	0,9573	1	0,9668
Y_{8ij}	-0,0081	0,5872	0,8135	0,8971	0,933	0,9501	0,9613	0,9668	1



Si en el modelo 2 el porcentaje de varianza de la pendiente suponía un 5% de la varianza total del modelo, ahora al incrementar dicha varianza a un 35%, a excepción de las correlaciones entre la primera ocasión de medida y el resto de mediciones que se mantienen nulas (valores resaltados en negrita en la primera medición), ya que, no se ha fijado varianza en el estatus. Se observa que se ha producido un incremento de todas las correlaciones respecto al modelo anterior (modelo 2). Así, por ejemplo, en el modelo analizado (modelo 2.1) la correlación entre la ocasión dos y tres es de 0,491, mientras que en el modelo anterior (modelo 2) era de 0,094. Por lo tanto, a medida que disminuyen las diferencias entre el porcentaje de varianza fijado en el error intrasujeto y el fijado en la varianza de la pendiente, incrementándose esta última varianza, se produce un aumento notable del valor de las correlaciones y, por ende, de la precisión con la que se estiman las puntuaciones de los estudiantes.

El incremento de dichos valores de correlación supone que ya no se requieran tantas mediciones, como en el modelo anterior, para la obtención de estimaciones fiables del rasgo de los estudiantes. Así, mientras que en el modelo 2 ni con nueve ocasiones de medidas se obtienen correlaciones muy altas (la correlación era de 0,739), que indiquen una buena fiabilidad en las estimaciones de las puntuaciones de los estudiantes; en el modelo 2.1, con ese mismo número de mediciones la correlación entre la ocasión ocho y nueve es casi perfecta (0,966). Incluso con un número menor de mediciones, como son cinco o seis, también se puede predecir los niveles de logro de los estudiantes en los cursos siguientes con precisión, ya que, las correlaciones siguen siendo superiores a 0,90.

En el gráfico de la tabla anterior (Tabla VI.7), se observa de manera más visual los valores de correlación que aparecen en la matriz de datos. En el caso de las correlaciones entre la ocasión uno y el resto de ocasiones, al ser valores que no se encuentran comprendidos entre los valores de la correlación (de 1 a -1), no se representan. El resto de correlaciones son de color azul, lo que refleja una direccionalidad positiva entre las mismas, además dicho color se vuelve cada vez más intenso a medida que se incrementa el número de mediciones, lo que indica que el valor de correlación va creciendo. Los círculos más pequeños se encuentran en las correlaciones entre la ocasión dos y el resto de ocasiones, ya que son correlaciones moderadas, comprendidas entre 0,50-0,60 (Wonnacott & Wonnacott, 1999)

Finalmente, dentro de esta subsituación 2 se presentan los resultados del **modelo 2.2**, en el cual, la varianza total del modelo se encuentra repartida equitativamente entre la varianza de la pendiente del nivel dos y la varianza del error intrasujeto.

Tabla VI.8. *Matriz de correlaciones del modelo 2.2*

MODELO 2.2

EFECTOS FIJOS		
Coef.		
β_{000}	**4,6**	
β_{100}	**0,4**	
EFECTOS ALEATORIOS		
NIVEL 3	$\sigma(u_{00j})$	$\sigma(u_{10j})$
$\sigma(u_{00j})$	0	
$\sigma(u_{10j})$		0
NIVEL 2	$\sigma(r_{0ij})$	$\sigma(r_{1ij})$
$\sigma(r_{0ij})$	0	
$\sigma(r_{1ij})$		**50%**
NIVEL 1	$\sigma(e_{tij})$	
$\sigma(e_{tij})$	**50%**	
R_{aulas}	0	
$R_{estudiantes}$	0	

MATRIZ DE CORRELACIONES

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	-0,0195	-0,009	-0,005	-0,0066	-0,0067	-0,0065	-0,0092	-0,0081
Y_{1ij}	-0,0195	1	**0,6347**	0,6715	0,6908	0,694	0,7004	0,7022	0,7049
Y_{2ij}	-0,009	0,6347	1	0,8466	0,8652	0,8749	0,8808	0,8837	0,887
Y_{3ij}	-0,005	0,6715	0,8466	1	**0,9195**	0,9286	0,9354	0,9381	0,9412
Y_{4ij}	-0,0066	0,6908	0,8652	0,9195	1	0,9509	0,9573	0,9603	0,9626
Y_{5ij}	-0,0067	0,694	0,8749	0,9286	0,9509	1	0,9666	0,9696	0,9724
Y_{6ij}	-0,0065	0,7004	0,8808	0,9354	0,9573	0,9666	1	0,9765	0,9788
Y_{7ij}	-0,0092	0,7022	0,8837	0,9381	0,9603	0,9696	0,9765	1	**0,9819**
Y_{8ij}	-0,0081	0,7049	0,887	0,9412	0,9626	0,9724	0,9788	0,9819	1

En la tabla anterior VI.8, sobre todo los valores resaltados, muestran que, a partir de la correlación entre la segunda y la tercera ocasión de medida, las correlaciones van desde una fiabilidad moderada (0,60) a muy alta (>0,90) (Wonnacott & Wonnacott, 1999), mostrando dicho modelo, valores superiores a los del modelo 2 y 2.1. Por ejemplo, en el modelo 2 la correlación entre la ocasión dos y tres era de 0,094 y en el modelo 2 de 0,491, mientras que en el modelo 2.2 es de 0,634.

Esta diferencia en los valores de correlación es menos pronunciada cuando el número de mediciones es elevado, sobre todo entre el modelo 2.1 y 2.2, pues mientras que en este último modelo con nueve ocasiones de medida la correlación entre la ocasión ocho y nueve es de 0,981, en el modelo 2.1 era de 0,966. La diferencia fundamental entre ambos modelos, es que mientras que en el modelo 2.1 se requieren entre cinco y seis mediciones para obtener correlaciones superiores a 0,90, en el modelo 2.2 se requieren entre cuatro y cinco medidas.

En el caso del modelo 2, como el porcentaje de varianza fijado en la pendiente era mucho más pequeño que el fijado en el modelo 2.1 y 2.2, se requiere de un mayor número de mediciones para obtener correlaciones muy altas.

Por último, hay que destacar también del modelo 2.2 que no hay mucha diferencia entre la correlación que se obtiene al realizar seis o siete mediciones y la obtenida con nueve. Por lo tanto, en este caso quizás no sea necesario realizar tantas aplicaciones como nueve, ya que, no aporta mucha información extra a la obtenida con siete medidas, pero si supone un coste mucho más elevado.

A la vista de los resultados obtenidos en los modelos de la **subsituación 2**, se observa que cuando los estudiantes tienen distinto rendimiento a lo largo del tiempo, debido a sus características sistemáticas y no al efecto de las características de las aulas, a medida que se incrementa el porcentaje de varianza fijado en la pendiente del nivel dos, acercándose al porcentaje de varianza fijado en el error intrasujeto, las correlaciones entre las ocasiones de medida son cada vez mayores, y, por lo tanto, más fiables con un menor número de mediciones.

Cuando en un determinado momento temporal se alcanza un valor de correlación entre dos ocasiones de medida muy alto, dicho valor se mantiene constante, y, por muchas mediciones extra que se realicen, el valor de correlación no se incrementa significativamente.

En cuanto al ERM, en estos modelos donde no se fija varianza en el estatus del nivel de estudiante ni del nivel de aula, entre la primera ocasión de medida y el resto de ocasiones se observa un patrón constante, en torno a 0, por lo tanto, el efecto de regresión a la media se mantiene inmutable a lo largo de todas las ocasiones de medida.

En las tablas siguientes (VI.9, VI.10 y VI.11) se muestran los resultados de los modelos de la **subsituación 3**, donde hay que recordar que el porcentaje fijado en la varianza de la pendiente del nivel de estudiante, era superior al fijado en la varianza del error intrasujeto.

La siguiente matriz de correlaciones corresponde al **modelo 3,** en el cual se fijan los mismos porcentajes de varianza que en el modelo 2, pero intercambiados entre la varianza de la pendiente del nivel dos y la varianza del error intrasujeto.

Tabla VI.9. *Matriz de correlaciones del modelo 3*

MODELO 3		
EFECTOS FIJOS		
Coef.		
β_{000}	**4,6**	
β_{100}	**0,4**	
EFECTOS ALEATORIOS		
NIVEL 3	$\sigma(u_{00j})$	$\sigma(u_{10j})$
$\sigma(u_{00j})$	0	
$\sigma(u_{10j})$		0
NIVEL 2	$\sigma(r_{0ij})$	$\sigma(r_{1ij})$
$\sigma(r_{0ij})$	0	
$\sigma(r_{1ij})$		**95%**
NIVEL 1	$\sigma(e_{tij})$	
$\sigma(e_{tij})$	**5%**	
R_{aulas}	0	
$R_{estudiantes}$	0	

MATRIZ DE CORRELACIONES

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**-0,0123**	**-0,0086**	**-0,0075**	**-0,0079**	**-0,0079**	**-0,0078**	**-0,0085**	**-0,0082**
Y_{1ij}	-0,0123	1	0,9687	0,972	0,9736	0,9738	0,9743	0,9744	0,9746
Y_{2ij}	-0,0086	0,9687	1	**0,9905**	0,9918	0,9924	0,9927	0,9929	0,9931
Y_{3ij}	-0,0075	0,972	0,9905	1	0,9954	0,996	0,9963	0,9965	0,9967
Y_{4ij}	-0,0079	0,9736	0,9918	0,9954	1	0,9973	0,9977	0,9978	0,998
Y_{5ij}	-0,0079	0,9738	0,9924	0,996	0,9973	1	0,9982	0,9984	0,9985
Y_{6ij}	-0,0078	0,9743	0,9927	0,9963	0,9977	0,9982	1	0,9987	0,9989
Y_{7ij}	-0,0085	0,9744	0,9929	0,9965	0,9978	0,9984	0,9987	1	**0,999**
Y_{8ij}	-0,0082	0,9746	0,9931	0,9967	0,998	0,9985	0,9989	0,999	1

Como se puede apreciar en la tabla anterior VI.9, en los valores de la primera medición resaltados en negrita, se sigue manteniendo un patrón constante y nulo entre la primera medición y el resto de ocasiones, debido a que, en general, tanto los estudiantes, como las aulas entre sí, parten de un mismo nivel de rendimiento.

Respecto al resto de correlaciones, son muy altas, ya que, se obtienen valores superiores a 0,90 (Wonnacott & Wonnacott, 1999). Esto es debido a que el porcentaje de varianza fijado en el error intrasujeto (5%) es muy pequeño en comparación con el porcentaje fijado en la varianza de la pendiente (95%). Bajo estas condiciones, con tres ocasiones de medida se puede tener una gran certeza del rendimiento real del estudiante, pues el valor de correlación entre la ocasión dos y tres es muy alto (0,969).

En este modelo, no existe mucha diferencia en la fiabilidad de las estimaciones de los niveles de logro de los estudiantes, cuando se realizan tres mediciones a cuando se realizan nueve, ya que, tal y como se puede observar en los valores destacados en la tabla, en ambas ocasiones se obtiene una información muy parecida. Sin embargo, en términos de coste la diferencia entre realizar tres o nueve mediciones es muy notable.

Mediante los resultados de dicho modelo, queda patente como el porcentaje de varianza fijado en la pendiente respecto al fijado en el error intrasujeto, afecta en la precisión de las estimaciones de las puntuaciones de los estudiantes. Cuanto mayor sea el porcentaje de varianza fijado en la pendiente, respecto a la varianza del error intrasujeto, se requiere de un menor número de mediciones para poder predecir el rendimiento académico de los estudiantes.

Si se comparan los resultados de este modelo con los del modelo 2, donde la varianza de la pendiente se fijaba a un 5% y la del error intrasujeto a un 95%, se observan unos valores de correlación muy distintos, ya que, en este modelo 2 con dos ocasiones de medida la correlación era prácticamente nula (0,094) y con nueve ocasiones era moderada (0,739) (Wonnacott & Wonnacott, 1999).

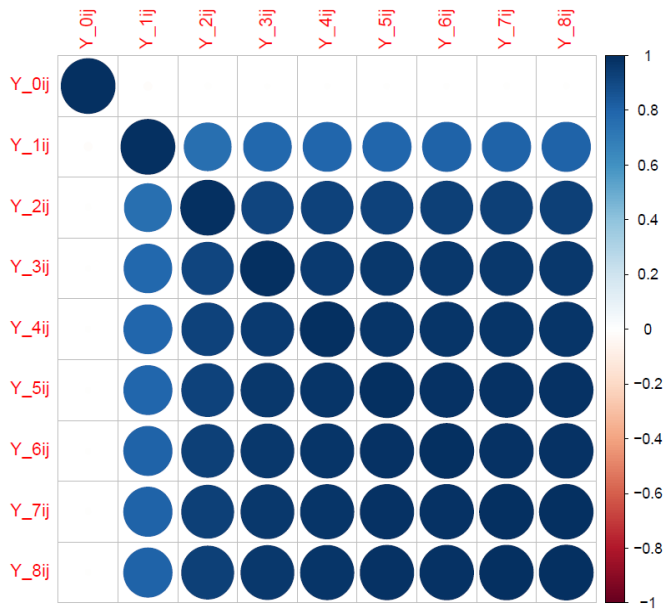
En la tabla siguiente VI.10 se presentan los resultados del **modelo 3.1,** donde no existe tanta diferencia entre el porcentaje de varianza fijado en la pendiente y el fijado en el error intrasujeto, en comparación con el modelo anterior.

Tabla VI.10. *Matriz de correlaciones del modelo 3.1*

MODELO 3.1

EFECTOS FIJOS			MATRIZ DE CORRELACIONES
Coef.			
β_{000}	**4,6**		
β_{100}	**0,4**		
EFECTOS ALEATORIOS			
NIVEL 3	$\sigma(u_{00j})$	$\sigma(u_{10j})$	
$\sigma(u_{00j})$	0		
$\sigma(u_{10j})$		0	
NIVEL 2	$\sigma(r_{0ij})$	$\sigma(r_{1ij})$	
$\sigma(r_{0ij})$	0		
$\sigma(r_{1ij})$		**65%**	
NIVEL 1	$\sigma(e_{tij})$		
$\sigma(e_{tij})$	**35%**		
R_{aulas}	0		
$R_{estudiantes}$	0		

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	-0,0181	-0,009	-0,0059	-0,007	-0,0072	-0,0069	-0,009	-0,0082
Y_{1ij}	-0,0181	1	**0,7588**	0,7839	0,7964	0,7984	0,8024	0,8035	0,8052
Y_{2ij}	-0,009	0,7588	1	0,9109	0,9221	0,9277	0,9311	0,9328	0,9347
Y_{3ij}	-0,0059	0,7839	0,9109	1	0,955	0,9602	0,964	0,9655	0,9672
Y_{4ij}	-0,007	0,7964	0,9221	0,955	1	0,973	0,9765	0,9782	0,9795
Y_{5ij}	-0,0072	0,7984	0,9277	0,9602	0,973	1	0,9818	0,9834	0,985
Y_{6ij}	-0,0069	0,8024	0,9311	0,964	0,9765	0,9818	1	0,9872	0,9885
Y_{7ij}	-0,009	0,8035	0,9328	0,9655	0,9782	0,9834	0,9872	1	**0,9901**
Y_{8ij}	-0,0082	0,8052	0,9347	0,9672	0,9795	0,985	0,9885	0,9901	1



Como se aprecia en los resultados de la tabla anterior V.10, cuando no existe tanta diferencia entre el porcentaje de varianza fijado en la pendiente y en el error

intrasujeto, los valores de la matriz de correlación obtenidos disminuyen respecto al modelo 3. Aunque dichas correlaciones siguen siendo superiores a las que se obtenían cuando el porcentaje de varianza del error intrasujeto era superior (65%) al fijado en la varianza del error intrasujeto (35%) (modelo 2.1), especialmente cuando el número de mediciones es bajo.

Mientras que en el modelo 2.1 la correlación entre la ocasión dos y tres era de 0,491, en el modelo analizado es de 0,758, tal y como se resalta en la tabla. Sin embargo, con nueve ocasiones la diferencia no es tan marcada debido al patrón acumulativo. En el modelo 2.1 la correlación entre la ocasión ocho y nueve era de 0,966 y en el modelo 3.1 es de 0,990, también en el modelo 3, el valor de correlación entre dichas ocasiones es similar a la de los modelos anteriores (0,999), a pesar de que el porcentaje fijado en la varianza de la pendiente era muy superior al que se había fijado en el modelo 2.1 y 3.1

En el modelo 3.1, con cuatro mediciones ya se obtienen correlaciones muy altas, superiores a 0,90. Por ello en el gráfico se observa que, a partir de la segunda medición, los círculos son grandes de color azul oscuro, mientras que en el modelo 2.3 era solo a partir de la ocasión tres cuando se apreciaban dichos círculos.

Finalmente, dentro de esta situación, se muestran los resultados del **modelo 3.2**, en el cual toda la varianza se había fijado en la varianza de la pendiente del nivel de estudiante.

Tabla VI.11. *Matriz de correlaciones del modelo 3.2*

MODELO 3.2		
EFECTOS FIJOS		
Coef.		
$\boldsymbol{\beta_{000}}$	**4,6**	
$\boldsymbol{\beta_{100}}$	**0,4**	
EFECTOS ALEATORIOS		
NIVEL 3	$\sigma(u_{00j})$	$\sigma(u_{10j})$
$\boldsymbol{\sigma(u_{00j})}$	0	
$\boldsymbol{\sigma(u_{10j})}$		0
NIVEL 2	$\sigma(r_{0ij})$	$\sigma(r_{1ij})$
$\boldsymbol{\sigma(r_{0ij})}$	0	
$\boldsymbol{\sigma(r_{1ij})}$		**100%**
NIVEL 1	$\sigma(e_{tij})$	
$\boldsymbol{\sigma(e_{tij})}$	0	
$\boldsymbol{R_{aulas}}$	0	
$\boldsymbol{R_{estudiantes}}$	0	

MATRIZ DE CORRELACIONES

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	**1**	NA	NA	NA	NA	NA	NA	NA	NA
Y_{1ij}	NA	**1**	1	1	1	1	1	1	1
Y_{2ij}	NA	1	**1**	1	1	1	1	1	1
Y_{3ij}	NA	1	1	**1**	1	1	1	1	1
Y_{4ij}	NA	1	1	1	**1**	1	1	1	1
Y_{5ij}	NA	1	1	1	1	**1**	1	1	1
Y_{6ij}	NA	1	1	1	1	1	**1**	1	1
Y_{7ij}	NA	1	1	1	1	1	1	**1**	1
Y_{8ij}	NA	1	1	1	1	1	1	1	**1**

Cuando no existe varianza en el error intrasujeto y solo se fija variabilidad en el crecimiento académico de los estudiantes, se obtiene una matriz de correlación perfecta, ya que la puntuación que se estima es siempre la misma en las distintas ocasiones de medida, no hay variabilidad en el error intrasujeto, por eso la correlación es perfecta.

En cuanto a las correlaciones entre la ocasión uno y el resto de mediciones no se pueden estimar, ya que no existe variabilidad ni en el estatus del nivel tres ni en el del nivel dos, tanto las aulas como los estudiantes parten del mismo rendimiento inicial.

A partir de los resultados obtenidos en los modelos de la subsituación 3, se puede concluir que cuando la varianza de la pendiente del nivel de estudiante es superior a la del error intrasujeto, se obtienen estimaciones más fiables que las obtenidas en la situación contraria (modelos de la subsituación 2). Además, a través de los modelos estudiados, se comprueba que cuanto mayor sea el porcentaje de varianza fijado en la pendiente, respecto al fijado en el error intrasujeto, se requieren menos mediciones para obtener estimaciones fiables de las puntuaciones de los estudiantes.

Bajo determinadas situaciones donde el porcentaje de varianza de la pendiente es muy elevado, entorno al 95%, en comparación con el fijado en la varianza del error intrsujeto (5%) (modelo 3.1), no existe una diferencia muy marcada en la precisión con la que se estiman las puntuaciones de los estudiantes, cuando se realizan pocas mediciones (tres y cuatro) que cuando se realizan muchas (ocho y nueve), ya que, se parten de valores de correlación muy altos que se mantienen constantes a lo largo del tiempo.

Los resultados de los modelos (Tabla VI.12, Tabla VI.13 y Tabla VI.14) que se presentan a continuación, se refieren a la **subsituación 4**, en la cual, se fijó varianza en el error intrasujeto y en el estatus del nivel 2.

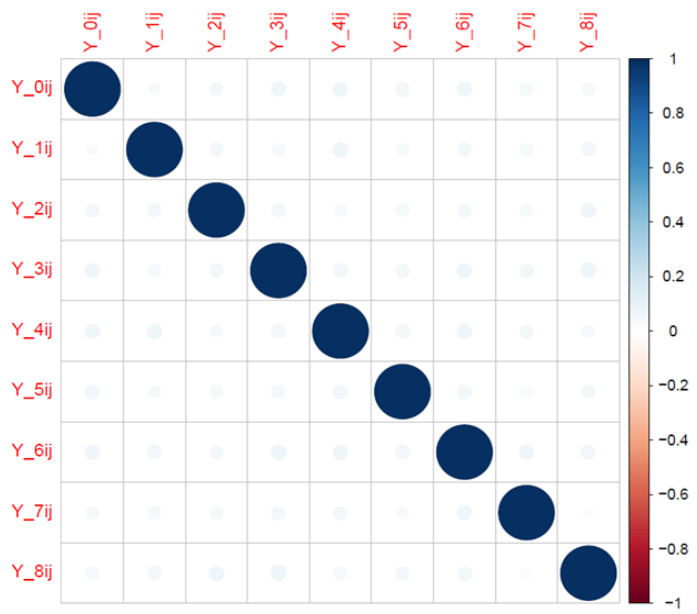
En el **modelo 4**, la proporción de varianza del error intrasujeto se fijó a un 95% de la varianza total y, el porcentaje restante se fijó en la varianza del intercepto.

Tabla VI.12. *Matriz de correlaciones del modelo 4*

MODELO 4		
EFECTOS FIJOS		
Coef.		
β_{000}	**4,6**	
β_{100}	**0,4**	
EFECTOS ALEATORIOS		
NIVEL 3	$\sigma(u_{00j})$	$\sigma(u_{10j})$
$\sigma(u_{00j})$	0	
$\sigma(u_{10j})$		0
NIVEL 2	$\sigma(r_{0ij})$	$\sigma(r_{1ij})$
$\sigma(r_{0ij})$	**5%**	
$\sigma(r_{1ij})$		0
NIVEL 1	$\sigma(e_{tij})$	
$\sigma(e_{tij})$	**95%**	
R_{aulas}	0	
$R_{estudiantes}$	0	

MATRIZ DE CORRELACIONES

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,0329**	**0,0504**	**0,0618**	**0,0605**	**0,0596**	**0,0618**	**0,0449**	**0,05**
Y_{1ij}	0,0329	1	**0,0542**	0,0471	0,0655	0,0408	0,0545	0,0466	0,0545
Y_{2ij}	0,0504	0,0542	1	0,0544	0,0428	0,0484	0,0506	0,0434	0,0607
Y_{3ij}	0,0618	0,0471	0,0544	1	0,0562	0,0517	0,0663	0,052	0,0696
Y_{4ij}	0,0605	0,0655	0,0428	0,0562	1	0,0566	0,0636	0,0515	0,048
Y_{5ij}	0,0596	0,0408	0,0484	0,0517	0,0566	1	0,0563	0,0369	0,0509
Y_{6ij}	0,0618	0,0545	0,0506	0,0663	0,0636	0,0563	1	0,0627	0,0578
Y_{7ij}	0,0449	0,0466	0,0434	0,052	0,0515	0,0369	0,0627	1	**0,0269**
Y_{8ij}	0,05	0,0545	0,0607	0,0696	0,048	0,0509	0,0578	0,0269	1



Cuando los estudiantes parten de distintos niveles de rendimiento previo (se fija varianza en el estatus del nivel de estudiante), tal y como se puede observar en los valores resaltados, se produce un ligero incremento en las correlaciones entre la primera ocasión de medida y el resto de ocasiones y un cambio de direccionalidad de las mismas, respecto a las correlaciones obtenidas en los modelos donde todos los estudiantes y aulas partían del mismo nivel de logro promedio (modelos de la subsituación 1, 2 y 3). Aun así, en este modelo dichas correlaciones siguen mostrando un patrón constante, en torno a 0, ya que el porcentaje de varianza fijado en el estatus es muy pequeño en comparación con la varianza fijada en el error intrasujeto. En este modelo el ERM se mantiene estable en todas las ocasiones de medida, ya que el patrón que se obtiene es constante (Nesselroade et al., 1980).

Que el valor del resto de correlaciones también sea tan bajo, tal y como se puede observar tanto en el gráfico como en la Tabla VI.12, donde incluso con nueve ocasiones de medida el valor de la correlación es nulo (0,027), se debe a que, además de que el porcentaje de varianza fijado en el estatus es muy pequeño, no se ha fijado variabilidad en la pendiente de ninguno de los dos niveles. Por ello, si se compara este modelo, con el modelo 2 donde no se había fijado variabilidad en el estatus de ninguno de los dos niveles (aula y estudiante), pero si en la varianza de la pendiente del nivel de estudiante (5%), las correlaciones son muy distintas, son muy superiores en el modelo 2, ya que se llega a obtener una correlación de 0,739 con nueve ocasiones de medida, a pesar de que la mayor parte de la varianza se fijaba en el error intrasujeto.

Por lo tanto, cuando los estudiantes parten de distinto rendimiento (modelo 4), pero su logro a lo largo del tiempo es constante, los valores de correlación entre la ocasión una y el resto de ocasiones son superiores a cuando todos los estudiantes tienen el mismo rendimiento inicial, pero este es distinto a lo largo del tiempo. Sin embargo, en lo que respecta al resto de correlaciones, se observa que cuando el rendimiento de los estudiantes a lo largo del tiempo es diferente (modelo 2), las correlaciones son superiores a las que se obtienen cuando no se fija varianza en la pendiente (modelo 4), esta diferencia es menos pronunciada cuando el número de ocasiones no es elevado. Así, mientras que, en el modelo 4 la correlación entre la ocasión dos y tres es de 0,054, en el modelo 2 era de 0,094; sin embargo, en este último modelo la correlación entre la ocasión ocho y nueve era de 0,739 y la del modelo 4 es de 0,026. Esta diferencia tan pronunciada cuando el número de medidas es elevado, se debe a que un valor distinto de cero en la varianza de la pendiente tiene un carácter acumulativo y, a medida que se incrementa el número de mediciones aumenta también el valor de las correlaciones. Sin embargo, si solo se fija varianza en el estatus y en el error intrasujeto, las correlaciones entre las puntuaciones de los estudiantes son muy similares tanto con tres medidas, como con nueve mediciones, ya que, no hay efecto de la pendiente de crecimiento y el valor de la varianza del estatus no es acumulativo.

En la siguiente Tabla VI.13 se muestran los resultados del **modelo 4.1,** en el cual se ha incrementado el porcentaje de varianza fijado en el estatus respecto al modelo anterior.

Tabla VI.13. *Matriz de correlaciones del modelo 4.1*

MODELO 4.1		
EFECTOS FIJOS		
Coef.		
β_{000}	**4,6**	
β_{100}	**0,4**	
EFECTOS ALEATORIOS		
NIVEL 3	$\sigma(u_{00j})$	$\sigma(u_{10j})$
$\sigma(u_{00j})$	0	
$\sigma(u_{10j})$		0
NIVEL 2	$\sigma(r_{0ij})$	$\sigma(r_{1ij})$
$\sigma(r_{0ij})$	**35%**	
$\sigma(r_{1ij})$		0
NIVEL 1	$\sigma(e_{tij})$	
$\sigma(e_{tij})$	**65%**	
R_{aulas}	0	
$R_{estudiantes}$	0	

MATRIZ DE CORRELACIONES

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,3396**	**0,3526**	**0,359**	**0,3613**	**0,3564**	**0,3585**	**0,3473**	**0,349**
Y_{1ij}	0,3396	1	0,3554	0,3492	0,365	0,3439	0,3537	0,3487	0,3523
Y_{2ij}	0,3526	0,3554	1	0,3552	0,3505	0,3501	0,352	0,3475	0,3576
Y_{3ij}	0,359	0,3492	0,3552	1	0,3583	0,351	0,3615	0,3521	0,3624
Y_{4ij}	0,3613	0,365	0,3505	0,3583	1	0,3575	0,3628	0,3549	0,3508
Y_{5ij}	0,3564	0,3439	0,3501	0,351	0,3575	1	0,3536	0,3407	0,3486
Y_{6ij}	0,3585	0,3537	0,352	0,3615	0,3628	0,3536	1	0,3588	0,3537
Y_{7ij}	0,3473	0,3487	0,3475	0,3521	0,3549	0,3407	0,3588	1	0,3329
Y_{8ij}	0,349	0,3523	0,3576	0,3624	0,3508	0,3486	0,3537	0,3329	1

Al incrementar el valor de varianza fijado en el intercepto y, por lo tanto, al no existir tanta diferencia entre el porcentaje fijado en dicha varianza y la varianza del error intrasujeto, las correlaciones se han incrementado bastante respecto al modelo 4.

Los valores de correlación que se obtienen en toda la matriz de datos son muy homogéneos, ya que no se fija variabilidad en la pendiente y, por lo tanto, no hay efecto del crecimiento. Respecto al patrón de correlación entre la primera ocasión de medida y el resto de ocasiones, se mantiene un patrón constante, pero con valores superiores a los del modelo 4.

Finalmente, dentro de esta subsituación, se presentan los resultados de un modelo (**modelo 4.2**), caracterizado porque existe la misma variabilidad en el error intrasujeto como en el estatus del nivel de estudiantes

Tabla VI.14. *Matriz de correlaciones del modelo 4.1*

MODELO 4.2		
EFECTOS FIJOS		
Coef.		
β_{000}	**4,6**	
β_{100}	**0,4**	
EFECTOS ALEATORIOS		
NIVEL 3	$\sigma(u_{00j})$	$\sigma(u_{10j})$
$\sigma(u_{00j})$	0	
$\sigma(u_{10j})$		0
NIVEL 2	$\sigma(r_{0ij})$	$\sigma(r_{1ij})$
$\sigma(r_{0ij})$	**50%**	
$\sigma(r_{1ij})$		0
NIVEL 1	$\sigma(e_{tij})$	
$\sigma(e_{tij})$	**50%**	
R_{aulas}	0	
$R_{estudiantes}$	0	

MATRIZ DE CORRELACIONES

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,4922**	**0,5018**	**0,5063**	**0,509**	**0,5039**	**0,5059**	**0,4976**	**0,4988**
Y_{1ij}	0,4922	1	0,5047	0,4994	0,5125	0,4949	0,5029	0,4993	0,502
Y_{2ij}	0,5018	0,5047	1	0,5037	0,5011	0,4993	0,5013	0,4981	0,5058
Y_{3ij}	0,5063	0,4994	0,5037	1	0,5065	0,4996	0,5081	0,5011	0,509
Y_{4ij}	0,509	0,5125	0,5011	0,5065	1	0,5055	0,5101	0,5043	0,5011
Y_{5ij}	0,5039	0,4949	0,4993	0,4996	0,5055	1	0,5016	0,4919	0,4979
Y_{6ij}	0,5059	0,5029	0,5013	0,5081	0,5101	0,5016	1	0,5063	0,5024
Y_{7ij}	0,4976	0,4993	0,4981	0,5011	0,5043	0,4919	0,5063	1	0,4866
Y_{8ij}	0,4988	0,502	0,5058	0,509	0,5011	0,4979	0,5024	0,4866	1

Como se puede observar en la Tabla VI.14 de resultados, al igualar los porcentajes de varianza fijados en el error intrasujeto y en el estatus del estudiante, se obtienen valores de correlación, en torno a 0,50. Estos valores son algo superiores a los del modelo 4.2, donde el porcentaje de varianza fijado en el intercepto era un poco menos del doble que el fijado en la varianza del error intrasujeto.

En conclusión, cuando los estudiantes parten de distinto rendimiento inicial, pero crecen al mismo ritmo, como es esperable, las correlaciones entre la ocasión uno y el resto de mediciones son superiores a las que se obtienen cuando los estudiantes parten del mismo nivel de logro, pero su crecimiento académico es distinto a lo largo del tiempo. Sin embargo, en esta última situación las correlaciones entre el resto de ocasiones de medida, son superiores a las que se obtiene cuando solo se fija varianza en el error intrasujeto y en el estatus. Esta diferencia entre los valores de correlación y, por tanto, en la fiabilidad de las estimaciones, es más pronunciada a medida que se incrementa el número de aplicaciones, debido al carácter acumulativo de la varianza de la pendiente. Además, se observa en los modelos de la subsituación 4, que los valores de correlación de toda la matriz de datos son muy homogéneos debido a la ausencia de crecimiento.

A continuación, se presentan los modelos referidos a la **subsituación 5** (Tabla VI. 15, Tabla VI. 16 y Tabla VI. 17). En estos modelos, se fijan los mismos porcentajes de varianza que en la subsituación 3, pero ahora referido a la varianza del estatus del nivel dos, en vez de a la varianza de la pendiente.

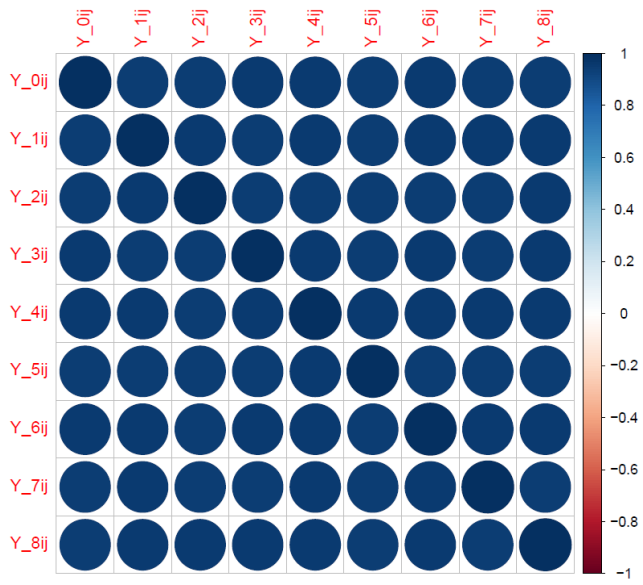
Tabla VI.15. *Matriz de correlaciones del modelo 5*

MODELO 5

EFECTOS FIJOS		
Coef.		
β_{000}	**4,6**	
β_{100}	**0,4**	
EFECTOS ALEATORIOS		
NIVEL 3	$\sigma(u_{00j})$	$\sigma(u_{10j})$
$\sigma(u_{00j})$	0	
$\sigma(u_{10j})$		0
NIVEL 2	$\sigma(r_{0ij})$	$\sigma(r_{1ij})$
$\sigma(r_{0ij})$	**95%**	
$\sigma(r_{1ij})$		0
NIVEL 1	$\sigma(e_{tij})$	
$\sigma(e_{tij})$	**5%**	
R_{aulas}	0	
$R_{estudiantes}$	0	

MATRIZ DE CORRELACIONES

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,9492**	0,9498	0,9501	0,9506	0,9497	0,9502	0,9495	0,9498
Y_{1ij}	0,9492	1	0,9504	0,9498	0,9513	0,9492	0,9502	0,95	0,9504
Y_{2ij}	0,9498	0,9504	1	0,9499	0,9498	0,9493	0,9498	0,9496	0,9505
Y_{3ij}	0,9501	0,9498	0,9499	1	0,9503	0,9492	0,9503	0,9497	0,9507
Y_{4ij}	0,9506	0,9513	0,9498	0,9503	1	0,95	0,9507	0,9502	0,9501
Y_{5ij}	0,9497	0,9492	0,9493	0,9492	0,95	1	0,9495	0,9486	0,9494
Y_{6ij}	0,9502	0,9502	0,9498	0,9503	0,9507	0,9495	1	0,9503	0,9501
Y_{7ij}	0,9495	0,95	0,9496	0,9497	0,9502	0,9486	0,9503	1	**0,9487**
Y_{8ij}	0,9498	0,9504	0,9505	0,9507	0,9501	0,9494	0,9501	0,9487	1



Como se observa en la Tabla anterior VI.15, al fijar un porcentaje de varianza muy pequeño en el error intrasujeto respecto a la varianza del estatus, se muestran valores de correlación muy diferentes a los obtenidos en el modelo contrario (modelo 4). En este modelo 5, tal y como también se puede observar en el gráfico, se obtienen correlaciones muy altas y homogéneas en toda la matriz de datos.

En el modelo 3, se fijó los mismos porcentajes de varianza, pero referidos a la varianza de la pendiente en vez de a la del estatus. En dicho modelo 3 todas las correlaciones eran mayores a 0,95 por lo que eran ligeramente superiores a las del modelo 5, a excepción de las correlaciones entre la ocasión una y el resto de ocasiones, que al no fijarse varianza en el estatus eran nulas y de direccionalidad negativa.

En ambos modelos (modelo 3 y 5), como el porcentaje fijado en el error intrasujeto es muy bajo, no existe mucha diferencia entre los valores de correlación obtenidos entre las puntuaciones de los estudiantes en las distintas ocasiones. Pero la interpretación de los mismos es muy distinta, ya que mientras que en el modelo 5 todos los estudiantes de una misma aula parten de distintos niveles de logro y crecen al mismo ritmo, en el modelo 3 los estudiantes de una misma clase cuentan con el mismo rendimiento inicial, pero su nivel de logro promedio es diferente a lo largo del tiempo.

En la siguiente tabla (VI.16) se presentan los resultados que se obtienen cuando el porcentaje de varianza del estatus sigue siendo superior al de la varianza del error intrasujeto, pero no existe tanta diferencia entre ambos valores (**modelo 5.1**)

Tabla VI.16. *Matriz de correlaciones del modelo 5.1*

MODELO 5.1		
EFECTOS FIJOS		
Coef.		
β_{000}	**4,6**	
β_{100}	**0,4**	
EFECTOS ALEATORIOS		
NIVEL 3	$\sigma(u_{00j})$	$\sigma(u_{10j})$
$\sigma(u_{00j})$	0	
$\sigma(u_{10j})$		0
NIVEL 2	$\sigma(r_{0ij})$	$\sigma(r_{1ij})$
$\sigma(r_{0ij})$	**65%**	
$\sigma(r_{1ij})$		0
NIVEL 1	$\sigma(e_{tij})$	
$\sigma(e_{tij})$	**35%**	
R_{aulas}	0	
$R_{estudiantes}$	0	

MATRIZ DE CORRELACIONES

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	0,**6445**	0,6508	0,6537	0,6562	0,6516	0,6535	0,6479	0,6489
Y_{1ij}	0,6445	1	0,6535	0,6495	0,6592	0,6459	0,652	0,6497	0,6518
Y_{2ij}	0,6508	0,6535	1	0,652	0,6508	0,6486	0,6505	0,6484	0,654
Y_{3ij}	0,6537	0,6495	0,652	1	0,6543	0,6485	0,6549	0,6502	0,6559
Y_{4ij}	0,6562	0,6592	0,6508	0,6543	1	0,6532	0,6569	0,653	0,6509
Y_{5ij}	0,6516	0,6459	0,6486	0,6485	0,6532	1	0,65	0,6434	0,6478
Y_{6ij}	0,6535	0,652	0,6505	0,6549	0,6569	0,65	1	0,654	0,6514
Y_{7ij}	0,6479	0,6497	0,6484	0,6502	0,653	0,6434	0,654	1	**0,6405**
Y_{8ij}	0,6489	0,6518	0,654	0,6559	0,6509	0,6478	0,6514	0,6405	1

En la tabla del modelo 5.1 se observa que cuando ya no existe tanta diferencia entre la varianza del error intrasujeto, y la varianza del estatus se obtienen valores muy inferiores a los del modelo 5.

En cuanto a los resultados generados con estos mismos porcentajes de varianza, pero invertidos entre la varianza del error intrasujeto y el estatus (modelo 4.1), son inferiores a los del modelo analizado (modelo 5.1), ya que, las correlaciones oscilaban en torno a 0,35.

Finalmente se presentan los resultados del **modelo 5.2** donde toda la varianza se fija en el estatus del nivel 2.

Tabla VI.17. *Matriz de correlaciones del modelo 5.2*

MODELO 5.2

EFECTOS FIJOS		
Coef.		
β_{000}	**4,6**	
β_{100}	**0,4**	
EFECTOS ALEATORIOS		
NIVEL 3	$\sigma(u_{00j})$	$\sigma(u_{10j})$
$\sigma(u_{00j})$	0	
$\sigma(u_{10j})$		0
NIVEL 2	$\sigma(r_{0ij})$	$\sigma(r_{1ij})$
$\sigma(r_{0ij})$	**100%**	
$\sigma(r_{1ij})$		0
NIVEL 1	$\sigma(e_{tij})$	
$\sigma(e_{tij})$	0	
R_{aulas}	0	
$R_{estudiantes}$	0	

MATRIZ DE CORRELACIONES

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	**1**	1	1	1	1	1	1	1	1
Y_{1ij}	1	**1**	1	1	1	1	1	1	1
Y_{2ij}	1	1	**1**	1	1	1	1	1	1
Y_{3ij}	1	1	1	**1**	1	1	1	1	1
Y_{4ij}	1	1	1	1	**1**	1	1	1	1
Y_{5ij}	1	1	1	1	1	**1**	1	1	1
Y_{6ij}	1	1	1	1	1	1	**1**	1	1
Y_{7ij}	1	1	1	1	1	1	1	**1**	1
Y_{8ij}	1	1	1	1	1	1	1	1	**1**

Como no se fija varianza en el error intrasujeto, pero si en el intercepto, toda la matriz de correlaciones que se obtiene es perfecta, ya que desde la ocasión una la puntuación que se estima es siempre la misma en todas las mediciones, el error que se comete en las diferentes mediciones es el mismo, debido a que no existe variabilidad.

Tras los resultados de los modelos de esta subsituación 5, se puede concluir algo similar a lo comentado para los modelos de la subsituación 2 y 3, y, es que, se obtienen valores de correlación más elevados cuando el porcentaje que se fija en el error intrasujeto es inferior al que se fija en la varianza del estatus.

Si se comparan los resultados de los modelos de la subsituación 2 y 3, donde los estudiantes parten del mismo rendimiento promedio pero crecen a distinto ritmo, con los de los modelos de la subsituación 4 y 5, donde los estudiantes tienen distintos niveles de logro inicial, pero su crecimiento es constante, se observa que en estos últimos modelos, se obtienen valores de correlación más altos entre la primera ocasión de medida y el resto de mediciones, que cuando solo se fija varianza en la pendiente. Sin embargo, el resto de valores generalmente son inferiores, ya que, no hay crecimiento, es como si solo se midiera una vez a todos los estudiantes y el valor de correlación obtenido entre la primera ocasión de medida con el resto de ocasiones, se mantuviera constante a lo largo del tiempo.

En las siguientes dos tablas (VI.18 y VI.19), se muestran los resultados de los modelos de la **subsituación 6,** en la cual se fijó variabilidad en el error intrasujeto y, en uno de los coeficientes aleatorios del nivel de estudiante.

En el **modelo 6,** que se presenta a continuación, se fijó los mismos porcentajes de varianza que en el modelo 2, pero en vez de referido a la pendiente del nivel de estudiante, referido a la pendiente del nivel de aula.

Tabla VI.18. *Matriz de correlaciones del modelo 6*

MODELO 6		
EFECTOS FIJOS		
Coef.		
$\boldsymbol{\beta_{000}}$	**4,6**	
$\boldsymbol{\beta_{100}}$	**0,4**	
EFECTOS ALEATORIOS		
NIVEL 3	$\sigma(u_{00j})$	$\sigma(u_{10j})$
$\boldsymbol{\sigma(u_{00j})}$	0	
$\boldsymbol{\sigma(u_{10j})}$		**5%**
NIVEL 2	$\sigma(r_{0ij})$	$\sigma(r_{1ij})$
$\boldsymbol{\sigma(r_{0ij})}$	0	
$\boldsymbol{\sigma(r_{1ij})}$		0
NIVEL 1	$\sigma(e_{tij})$	
$\boldsymbol{\sigma(e_{tij})}$	**95%**	
$\boldsymbol{R_{aulas}}$	0	
$\boldsymbol{R_{estudiantes}}$	0	

MATRIZ DE CORRELACIONES

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	-0,0188	-0,003	0,0077	0,0047	0,0049	0,0063	-0,0035	0,0006
Y_{1ij}	-0,0188	1	**0,0967**	0,1235	0,1624	0,1612	0,1847	0,1898	0,202
Y_{2ij}	-0,003	0,0967	1	0,2364	0,2746	0,3087	0,3365	0,3507	0,373
Y_{3ij}	0,0077	0,1235	0,2364	1	0,3824	0,419	0,4616	0,4781	0,5037
Y_{4ij}	0,0047	0,1624	0,2746	0,3824	1	0,5054	0,5505	0,5735	0,5941
Y_{5ij}	0,0049	0,1612	0,3087	0,419	0,5054	1	0,6025	0,6264	0,6555
Y_{6ij}	0,0063	0,1847	0,3365	0,4616	0,5505	0,6025	1	0,6879	0,7117
Y_{7ij}	-0,0035	0,1898	0,3507	0,4781	0,5735	0,6264	0,6879	1	0**,7401**
Y_{8ij}	0,0006	0,202	0,373	0,5037	0,5941	0,6555	0,7117	0,7401	1

Como se aprecia en la tabla anterior V.18, si se comparan los resultados de este modelo 6, con los del modelo 2, en ambos modelos se generan valores de correlación muy similares. Por ejemplo, en el modelo 2, la correlación entre la ocasión dos y tres era de 0,094 y la correlación entre la ocasión ocho y nueve era de 0,739; en el modelo 6, la correlación entre la ocasión dos y tres es de 0,096 y la de la ocasión ocho y nueve de 0,740. De acuerdo con estos resultados, parece no haber a penas diferencia en la matriz de correlaciones entre ambos modelos, y, por lo tanto, en la fiabilidad de las estimaciones en las puntuaciones de los estudiantes, cuando se fija el mismo porcentaje en la varianza de la pendiente del nivel de estudiante que en la varianza de la pendiente del nivel de aula y, por supuesto, los mismos valores en la varianza del error intrasujeto de ambos modelos.

En el siguiente modelo, **modelo 6.1**, se fijaron los mismos porcentajes de varianza que en el modelo 4, pero en vez de fijar a un 5% la varianza del estatus del nivel de estudiante, se fijó la varianza del intercepto del nivel tres.

Tabla VI.19. *Matriz de correlaciones del modelo 6.1*

MODELO 6.1

EFECTOS FIJOS		
Coef.		
$\boldsymbol{\beta_{000}}$	**4,6**	
$\boldsymbol{\beta_{100}}$	**0,4**	
EFECTOS ALEATORIOS		
NIVEL 3	$\sigma(u_{00j})$	$\sigma(u_{10j})$
$\boldsymbol{\sigma(u_{00j})}$	**5%**	
$\boldsymbol{\sigma(u_{10j})}$		0
NIVEL 2	$\sigma(r_{0ij})$	$\sigma(r_{1ij})$
$\boldsymbol{\sigma(r_{0ij})}$	0	
$\boldsymbol{\sigma(r_{1ij})}$		0
NIVEL 1	$\sigma(e_{tij})$	
$\boldsymbol{\sigma(e_{tij})}$	**95%**	
$\boldsymbol{R_{aulas}}$	0	
$\boldsymbol{R_{estudiantes}}$	0	

MATRIZ DE CORRELACIONES

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,0304**	0,0448	0,0569	0,0546	0,0569	0,0581	0,0405	0,0451
Y_{1ij}	0,0304	1	0,0502	0,0437	0,0612	0,0398	0,0524	0,0438	0,0512
Y_{2ij}	0,0448	0,0502	1	0,0479	0,0352	0,0441	0,0452	0,0374	0,0543
Y_{3ij}	0,0569	0,0437	0,0479	1	0,0493	0,0481	0,0617	0,0467	0,0639
Y_{4ij}	0,0546	0,0612	0,0352	0,0493	1	0,052	0,058	0,0452	0,0411
Y_{5ij}	0,0569	0,0398	0,0441	0,0481	0,052	1	0,054	0,0338	0,0474
Y_{6ij}	0,0581	0,0524	0,0452	0,0617	0,058	0,054	1	0,0586	0,0532
Y_{7ij}	0,0405	0,0438	0,0374	0,0467	0,0452	0,0338	0,0586	1	**0,0216**
Y_{8ij}	0,0451	0,0512	0,0543	0,0639	0,0411	0,0474	0,0532	0,0216	1

Al comparar los resultados de dicho modelo, con el modelo 4, se confirma nuevamente, que el nivel en el que se fija varianza no afecta de manera significativa en la precisión con la que se estiman los niveles de logro de los estudiantes, siempre y cuando, el porcentaje de varianza fijado en este coeficiente y en el resto de los coeficientes sea el mismo. Así, por ejemplo, en el modelo 4, donde la varianza fijada en el estatus del nivel de estudiante era un 5% de la varianza total, la correlación entre la ocasión una y dos era de 0,032, muy similar a la del modelo 6.1 que es de

0,030, estas semejanzas entre los valores de correlación de ambos modelos se mantienen en el resto de valores de la matriz.

A la vista de los resultados de los modelos de la subsituación 6, parece que el nivel (estudiante o aula) donde se fija varianza, no afecta en la fiabilidad con la que se estiman las puntuaciones de los estudiantes, ya que, cuando se comparan modelos donde se fija los mismos porcentajes de varianza en los mismos coeficientes, y sólo cambia el nivel donde se fija la varianza del intercepto o de la pendiente, los resultados de la matriz de correlaciones son muy similares. Sin embargo, la interpretación de ambos modelos es muy distinta, ya que al fijar variabilidad solo en el nivel de estudiantes el rendimiento se ve afectado por las características de los mismos, mientras que si sólo se fija en el nivel tres se debe al efecto de las aulas, por ende, resultados similares pueden deberse a modelos muy diferentes.

A continuación, en los modelos de la subsituación 7 y 8, se muestra los resultados de modelos donde además de fijar variabilidad en el error intrasujeto, se fija cierto porcentaje de varianza en el estatus y en la pendiente de un mismo nivel.

En los modelos de la **subsituación 7** (Tabla VI.20, Tabla VI.21, Tabla VI.22 y Tabla VI.23) que se presentan a continuación, el porcentaje de varianza fijado en el error intrasujeto, es superior o igual a la suma de los porcentajes de varianzas fijados en el estatus y en la pendiente de un mismo nivel.

En el **modelo 7** los estudiantes partían de distintos niveles de rendimiento previo (se fijó varianza en el intercepto de los estudiantes (2,5%)) y, su rendimiento a lo largo del tiempo era distinto (se fijó varianza en la pendiente del nivel dos (2,5%)). Los porcentajes de varianza totales fijados en cada uno de los niveles (5% en el nivel de estudiante y 95% en la varianza del error intrasujeto), son los mismos porcentajes totales que los fijados en el modelo 2 y 4, pero difieren entre sí, en cómo se ha repartido la varianza dentro del nivel 2.

Tabla VI.20. *Matriz de correlaciones del modelo 7*

MODELO 7		
EFECTOS FIJOS		
Coef.		
β_{000}	**4,6**	
β_{100}	**0,4**	
EFECTOS ALEATORIOS		
NIVEL 3	$\sigma(u_{00j})$	$\sigma(u_{10j})$
$\sigma(u_{00j})$	0	
$\sigma(u_{10j})$		0
NIVEL 2	$\sigma(r_{0ij})$	$\sigma(r_{1ij})$
$\sigma(r_{0ij})$	**2,5%**	
$\sigma(r_{1ij})$		**2,5%**
NIVEL 1	$\sigma(e_{tij})$	
$\sigma(e_{tij})$	**95%**	
R_{aulas}	0	
$R_{estudiantes}$	0	

MATRIZ DE CORRELACIONES

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,0102**	**0,0233**	**0,0282**	**0,0271**	**0,0214**	**0,0216**	**0,0079**	**0,0112**
Y_{1ij}	0,0102	1	**0,0796**	0,0887	0,1233	0,1123	0,1343	0,1365	0,1487
Y_{2ij}	0,0233	0,0796	1	0,1489	0,1723	0,1981	0,2213	0,232	0,2562
Y_{3ij}	0,0282	0,0887	0,1489	1	0,2473	0,2761	0,3171	0,3302	0,3595
Y_{4ij}	0,0271	0,1233	0,1723	0,2473	1	0,3467	0,3899	0,4109	0,4331
Y_{5ij}	0,0214	0,1123	0,1981	0,2761	0,3467	1	0,4402	0,4629	0,4984
Y_{6ij}	0,0216	0,1343	0,2213	0,3171	0,3899	0,4402	1	0,5313	0,5597
Y_{7ij}	0,0079	0,1365	0,232	0,3302	0,4109	0,4629	0,5313	1	**0,5899**
Y_{8ij}	0,0112	0,1487	0,2562	0,3595	0,4331	0,4984	0,5597	0,5899	1

En los resultados del modelo 7, se observa resaltado en negrita, que las correlaciones entre la primera ocasión de medida y el resto de las ocasiones, proporcionan más información que si todos los estudiantes parten del mismo nivel de logro; aunque el valor de correlación es casi nulo, ya que el porcentaje de varianza fijado en el estatus es muy pequeño, en comparación con el porcentaje fijado en el error intrasujeto.

Por ello, si se comparan estos valores de correlación entre la ocasión una y el resto de las ocasiones del modelo 7, con el modelo 4, en el que la varianza del intercepto se fijó a un 5% de la varianza total del modelo. Este modelo 4 muestra valores de correlación superiores al modelo 7, siendo el porcentaje de varianza fijado en el error intrasujeto el mismo en ambos modelos. Así, por ejemplo, mientras que en el modelo 4 la correlación entre la ocasión una y dos era de 0,329, en el modelo 7 es de 0,010.

En cuanto al resto de correlaciones, se obtienen valores superiores en este modelo 7, que se fija varianza tanto en el intercepto como en la pendiente del nivel de estudiante, que en el modelo 4, donde la varianza de la pendiente se fijó a 0.

Por el contrario, estos valores de correlación del modelo 7, son inferiores a los obtenidos en el modelo 2, donde solo se fijó varianza en la pendiente del nivel de estudiante (5%) y en el error intrasujeto. Muestra de ello es, por ejemplo, la correlación entre la ocasión ocho y nueve que en el modelo 2 era de 0,739 y en el

modelo 7, es de 0,589. Que a partir de la primera ocasión de medida los valores de correlación obtenidos en el modelo 2 sean superiores a los del modelo 7, se debe a que el porcentaje de varianza de la pendiente fijado en el modelo 2 es el doble que el que se fijó en el modelo 7.

En la tabla siguiente VI.21, se muestra la matriz de correlaciones del **modelo 7.1** donde el porcentaje de varianza se encuentra repartido, en un 50% entre la varianza del error intrasujeto y la varianza del nivel de estudiante, tal y como ocurría en los modelos 2.2 y 4.2.

Tabla VI.21. *Matriz de correlaciones del modelo 7.1*

MODELO 7.1		
EFECTOS FIJOS		
Coef.		
$\boldsymbol{\beta_{000}}$	**4,6**	
$\boldsymbol{\beta_{100}}$	**0,4**	
EFECTOS ALEATORIOS		
NIVEL 3	$\sigma(u_{00j})$	$\sigma(u_{10j})$
$\boldsymbol{\sigma(u_{00j})}$	0	
$\boldsymbol{\sigma(u_{10j})}$		0
NIVEL 2	$\sigma(r_{0ij})$	$\sigma(r_{1ij})$
$\boldsymbol{\sigma(r_{0ij})}$	**25%**	
$\boldsymbol{\sigma(r_{1ij})}$		**25%**
NIVEL 1	$\sigma(e_{tij})$	
$\boldsymbol{\sigma(e_{tij})}$	**50%**	
$\boldsymbol{R_{aulas}}$	0	
$\boldsymbol{R_{estudiantes}}$	0	

MATRIZ DE CORRELACIONES

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,2867**	**0,2246**	**0,1699**	**0,14**	**0,1112**	**0,0951**	**0,0796**	**0,071**
Y_{1ij}	0,2867	1	0,5772	0,5802	0,585	0,5688	0,5655	0,5585	0,5546
Y_{2ij}	0,2246	0,5772	1	0,7619	0,7786	0,782	0,7839	0,7826	0,7832
Y_{3ij}	0,1699	0,5802	0,7619	1	0,861	0,8706	0,8784	0,8797	0,8822
Y_{4ij}	0,14	0,585	0,7786	0,861	1	0,9097	0,9187	0,9221	0,9242
Y_{5ij}	0,1112	0,5688	0,782	0,8706	0,9097	1	0,9372	0,9416	0,9457
Y_{6ij}	0,0951	0,5655	0,7839	0,8784	0,9187	0,9372	1	0,9551	0,9587
Y_{7ij}	0,0796	0,5585	0,7826	0,8797	0,9221	0,9416	0,9551	1	**0,9649**
Y_{8ij}	0,071	0,5546	0,7832	0,8822	0,9242	0,9457	0,9587	0,9649	1

Lo primero que destaca del modelo 7.1, es que por primera vez se observa una cierta tendencia hacia un patrón decreciente entre la primera ocasión de medida y el resto de ocasiones (valores en negrita), lo que supondría que el ERM, se incrementa a medida que se añaden mas mediciones.

Además, tal y como se ha comentado para el modelo 7, la intensidad de los valores de correlación de las puntuaciones de los estudiantes en las distintas ocasiones de medida, depende del porcentaje de varianza que se fija en los coeficientes aleatorios. Así, por ejemplo, si se compara los resultados del modelo 7.1 con los del modelo 4.2, donde la varianza del estatus se fijó a un 35%, este modelo proporciona valores de correlación entre la ocasión una y el resto de las ocasiones superiores a los del modelo 7.1. Además, en este último modelo, al existir una cierta

tendencia hacia un patrón decreciente, a medida que se añaden mediciones esos valores se van reduciendo y la diferencia entre ambos modelos es más marcada.

En cuanto al resto de ocasiones, en dicho modelo 7.1, con cinco o seis mediciones las puntuaciones de los estudiantes se estiman con bastante precisión, mientras que en el modelo 4.1, al no fijarse varianza en la pendiente y el porcentaje de varianza fijado en el intercepto no ser muy alto, con nueve medidas la correlación era inferior a 0,35. Sin embargo, en el modelo 2.2, como se fijó un porcentaje de varianza del 35% en la pendiente del nivel de estudiante, a partir de la primera ocasión de medida, las correlaciones son superiores a las del modelo 7.1, aunque esta diferencia no es tan marcada cuando el número de mediciones es elevado . Así, por ejemplo, mientras que en el modelo 7.1 la correlación entre la ocasión ocho y nueve es de 0,965 en el modelo 2.2 era de 0,982

Posteriormente, dentro de la subsituación 7, se muestra el resultado de dos modelos donde el porcentaje de varianza que se fijó en la pendiente del nivel de estudiante, no es el mismo que el que se fijó en la varianza del estatus.

En la tabla contigua se muestra los resultados del **modelo 7.2,** donde el porcentaje fijado en la varianza del intercepto del nivel de estudiante, es superior al fijado en la varianza de la pendiente.

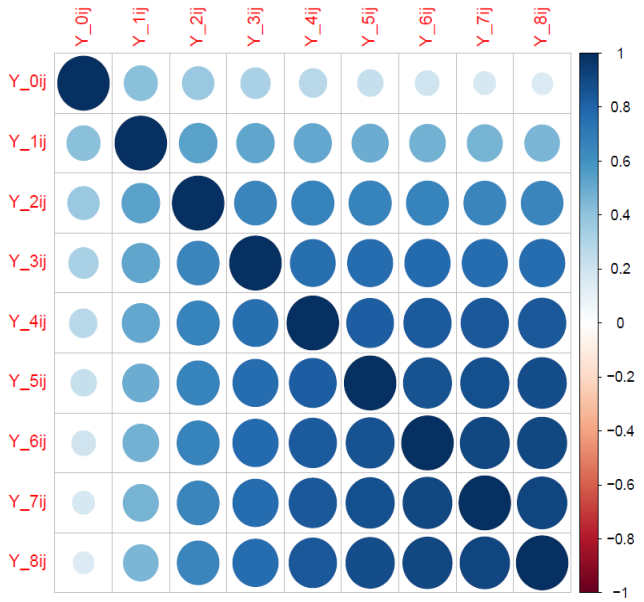
Tabla VI.22. *Matriz de correlaciones del modelo 7.2*

MODELO 7.2

EFECTOS FIJOS		
Coef.		
β_{000}	**4,6**	
β_{100}	**0,4**	
EFECTOS ALEATORIOS		
NIVEL 3	$\sigma(u_{00j})$	$\sigma(u_{10j})$
$\sigma(u_{00j})$	0	
$\sigma(u_{00j})$		0
NIVEL 2	$\sigma(r_{0ij})$	$\sigma(r_{1ij})$
$\sigma(r_{0ij})$	**40%**	
$\sigma(r_{1ij})$		**10%**
NIVEL 1	$\sigma(e_{tij})$	
$\sigma(e_{tij})$	**50%**	
R_{aulas}	0	
$R_{estudiantes}$	0	

MATRIZ DE CORRELACIONES

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,412**	**0,3703**	**0,3219**	**0,2709**	**0,233**	**0,202**	**0,1734**	**0,1541**
Y_{1ij}	0,412	1	**0,5332**	**0,5262**	**0,5182**	**0,4923**	**0,4792**	**0,4627**	**0,4507**
Y_{2ij}	0,3703	0,5332	1	0,6559	0,6674	0,6674	0,6639	0,6566	0,6527
Y_{3ij}	0,3219	0,5262	0,6559	1	0,7538	0,7626	0,7707	0,7686	0,7691
Y_{4ij}	0,2709	0,5182	0,6674	0,7538	1	0,8244	0,8382	0,8422	0,8439
Y_{5ij}	0,233	0,4923	0,6674	0,7626	0,8244	1	0,8669	0,8738	0,8802
Y_{6ij}	0,202	0,4792	0,6639	0,7707	0,8382	0,8669	1	0,9007	0,907
Y_{7ij}	0,1734	0,4627	0,6566	0,7686	0,8422	0,8738	0,9007	1	**0,9199**
Y_{8ij}	0,1541	0,4507	0,6527	0,7691	0,8439	0,8802	0,907	0,9199	1



En los resultados de la tabla anterior (VI.22), se muestra que al fijar un porcentaje de varianza en el estatus del nivel de estudiante superior al de algunos modelos anteriores, por ejemplo, al modelo 4 (la varianza del estatus se fijó a un 5%) o al modelo 4.1 (la varianza del estatus se fijó a un 35%), se obtienen valores de correlación entre la ocasión una y el resto de mediciones, superiores a todos los valores del modelo 4. Pero si se compara con el modelo 4.1, tan solo muestra valores superiores en las correlaciones entre la ocasión una con la ocasión dos y con la ocasión tres, ya que el patrón decreciente de dicho modelo 7.2, hace que según se realizan más mediciones disminuyan los valores de correlación, y, por tanto, que

estos valores sean inferiores a los del modelo 4.1, donde el porcentaje de varianza fijado en el intercepto, era inferior al fijado en el modelo 7.2

En cuanto a las correlaciones entre el resto de las ocasiones de medida, lo primero que se aprecia, y por ello se ha resaltado, es que, en las correlaciones de la segunda ocasión de medida con el resto de mediciones, sigue predominando una tendencia hacia un patrón decreciente. En dicho modelo, aunque la varianza de la pendiente se fijó a un 10% de la varianza total del modelo, con nueve ocasiones de medida se obtiene un valor de correlación superior a 0,90 y, con cinco superior a 0,80. Que se obtengan estos valores de correlación, siendo tan pequeño el porcentaje de varianza fijado en la pendiente (10%) se debe, a que, al fijar a un 40% la varianza en el estatus, se parten de valores de correlación muy altos, esto unido al carácter acumulativo de la varianza de la pendiente, da como resultado dichos valores de correlación.

Si se comparan los valores de correlación, a partir de la primera ocasión de medida, del modelo 7.2 con el modelo 2, donde se fijó la varianza de la pendiente a un 5%, se muestran valores de correlación superiores al modelo 2. Sin embargo, los valores del modelo 7.2, son muy próximos al modelo 2.1, en el cual la varianza de la pendiente se fijaba a un 35%, e inferiores al modelo 2.2, ya que, en este modelo la varianza de la pendiente se fijó a un 50%, aunque nuevamente dicha diferencia entre ambos modelos (modelo 2.2 y modelo 7.2), es más marcada cuando el número de mediciones es pequeño.

Los mismos porcentajes de varianza del modelo 7.2, fueron fijados en el **modelo 7.3**, pero referido al nivel de aula. Los resultados de dicho modelo se presentan en el Anexo III, ya que son muy similares a los obtenidos en el modelo 7.2, pues como se ha demostrado, el nivel donde se fija varianza no afecta a la precisión con la que se estiman las puntuaciones de los estudiantes.

Finalmente, dentro de esta subsituación se muestran los resultados del **modelo 7.4,** donde el porcentaje de varianza fijado en la pendiente fue del 40% y, el fijado en el estatus del 10%.

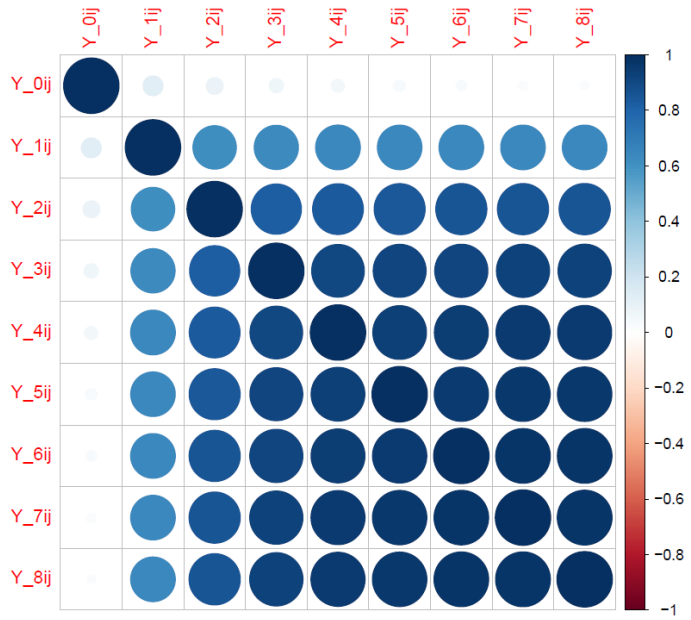
Tabla VI.23. *Matriz de correlaciones del modelo 7.4*

MODELO 7.4

EFECTOS FIJOS		
Coef.		
β_{000}	**4,6**	
β_{100}	**0,4**	
EFECTOS ALEATORIOS		
NIVEL 3	$\sigma(u_{00j})$	$\sigma(u_{10j})$
$\sigma(u_{00j})$	0	
$\sigma(u_{10j})$		0
NIVEL 2	$\sigma(r_{0ij})$	$\sigma(r_{1ij})$
$\sigma(r_{0ij})$	**10%**	
$\sigma(r_{1ij})$		**40%**
NIVEL 1	$\sigma(e_{tij})$	
$\sigma(e_{tij})$	**50%**	
R_{aulas}	0	
$R_{estudiantes}$	0	

MATRIZ DE CORRELACIONES

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,1227**	**0,0888**	**0,0636**	**0,0512**	**0,0388**	**0,0329**	**0,0255**	**0,0229**
Y_{1ij}	0,1227	1	0,6133	0,6358	0,65	0,6457	0,6487	0,6475	0,6479
Y_{2ij}	0,0888	0,6133	1	0,8204	0,8387	0,8461	0,8507	0,8523	0,8548
Y_{3ij}	0,0636	0,6358	0,8204	1	0,9033	0,9125	0,9196	0,9219	0,9248
Y_{4ij}	0,0512	0,65	0,8387	0,9033	1	0,9398	0,9469	0,95	0,9522
Y_{5ij}	0,0388	0,6457	0,8461	0,9125	0,9398	1	0,9589	0,9623	0,9655
Y_{6ij}	0,0329	0,6487	0,8507	0,9196	0,9469	0,9589	1	0,971	0,9736
Y_{7ij}	0,0255	0,6475	0,8523	0,9219	0,95	0,9623	0,971	1	0,9775
Y_{8ij}	0,0229	0,6479	0,8548	0,9248	0,9522	0,9655	0,9736	0,9775	1



Si se comparan los resultados de este modelo 7.4, con los resultados del modelo 7.2, se observa que los valores de correlación entre la primera ocasión de medida con el resto de las ocasiones, son inferiores en el modelo 7.4, ya que, el porcentaje de varianza fijado en el estatus es menor al del modelo 7.2. Además, en el modelo 7.2 entre las correlaciones de la primera ocasión de medida predominaba un patrón decreciente más marcado, que se extendía hasta las correlaciones entre la segunda ocasión con el resto de mediciones. Sin embargo, en el modelo 7.4 se observa un patrón más constante, debido a que, el porcentaje fijado en dicho coeficiente aleatorio, es muy bajo, si se compara con el resto de porcentajes fijados en la varianza de la pendiente y en el error intrasujeto.

En cuanto al resto de correlaciones, en este modelo son superiores a las del modelo 7.2, sobre todo cuando el número de mediciones es pequeño. Así, por ejemplo, mientras que en el modelo 7.2, la correlación entre la ocasión dos y tres era de 0,533, en el modelo 7.4, es de 0,613, pero con nueve ocasiones ambos modelos cuentan con correlaciones superiores a 0,90; a pesar de que el porcentaje de varianza fijado en el modelo 7.4 es tres veces superior al fijado en el modelo 7.2.

Dichas diferencias entre los valores de correlación del modelo 7.2 y del modelo 7.3, cuando el número de mediciones es pequeño, se aprecian de manera más visual en los gráficos correspondientes a cada una de las matrices de correlaciones. Así, mientras que, en el modelo 7.2, las correlaciones entre la ocasión una y el resto de las ocasiones, están representadas con círculos de color azul más intenso y con un tamaño que va decreciendo a medida que se añaden mediciones; en el modelo 7.4, al ser estas correlaciones nulas, apenas se aprecian. En cuanto al resto de correlaciones, cuando el número de mediciones es elevado, los círculos de ambos modelos son semejante en tamaño y color, pero cuando el número de mediciones es pequeño, los círculos del modelo 7.4 presentan colores más intensos de tamaños más grandes, que los del modelo 7.2

Tras los resultados de los modelos de la subsituación 7, se puede concluir, que la fiabilidad de las estimaciones de las puntuaciones de los estudiantes depende del porcentaje de varianza fijado en los coeficientes aleatorios, y de la diferencia existente entre los mismos.

Cuando se fija variabilidad tanto en el intercepto como en la pendiente de un mismo nivel, las correlaciones entre la primera ocasión de medida con el resto de las ocasiones son superiores a las que se obtienen en los modelos en los cuales solo se fija varianza en la pendiente, pero no en el estatus. Sin embargo, si se compara los resultados de un modelo donde se fija varianza tanto en la pendiente como en el intercepto (siendo el porcentaje de varianza fijado en el estutus ligeramente inferior al porcentaje fijado en el error intrasujeto (modelo 7.2)), con un modelo en el que se fija un porcentaje inferior en la varianza del intercepto respecto al modelo anterior (por ejemplo, modelo 4.1), es posible que en determinadas correlaciones entre la primera medida y el resto de ocasiones, este último modelo muestre valores de correlación iguales o superiores al del modelo donde el porcentaje de varianza fijado en el intercepto era mayor. Esto es debido al patrón decreciente que se genera entre

la primera ocasión de medida con el resto de las ocasiones, cuando se considera que los estudiantes parten de distintos niveles de logro y crecen a distinto ritmo, dicho patrón es mas marcado cuanto más elevado sea el porcentaje de varianza fijado en el estatus respecto a la varianza de la pendiente.

El patrón decreciente entre la primera ocasión de medida y el resto de ocasiones, indica, según Nesselroade et al., (1980), que el efecto de regresión a la media tiende a incrementarse a medida que se añaden mediciones.

En cuanto al resto de correlaciones, como es esperable, en los modelos donde se fija varianza en el intercepto y en la pendiente de un mismo nivel, a partir de la ocasión dos, muestran valores de correlación superiores a los que se obtienen en aquellos modelos donde el ritmo de crecimiento educativo a lo largo del tiempo es el mismo, aunque el rendimiento promedio previo es distinto, siempre y cuando el porcentaje de varianza fijado en el estatus no sea muy alto, en comparación con el porcentaje fijado en la pendiente del otro modelo. Ya que, los valores de correlación generados en la primera ocasión de medida, debido al porcentaje de varianza fijado en el estatus, se mantienen homogéneos a lo largo de toda la matriz de correlaciones.

Respecto a los modelos donde existe variabilidad sólo en la pendiente y en el error intrasujeto, si las correlaciones de este modelo se comparan con las que se obtienen en un modelo donde se fija varianza tanto en la pendiente como en el intercepto de un mismo nivel, son muy semejantes, siempre y cuando, la varianza del error intrasujeto fijada sea la misma. Esta semejanza es más marcada cuando el número de ocasiones de medidas son altas que cuando se realizan pocas mediciones; en este caso, con pocas ocasiones, son ligeramente superiores las correlaciones del modelo en el que se fija un mayor porcentaje de varianza en la pendiente.

Dentro de la **subsituación 8,** en las siguientes tablas (Tabla VI.24, Tabla VI.25 y Tabla VI.26) se presentan los resultados de los distintos modelos.

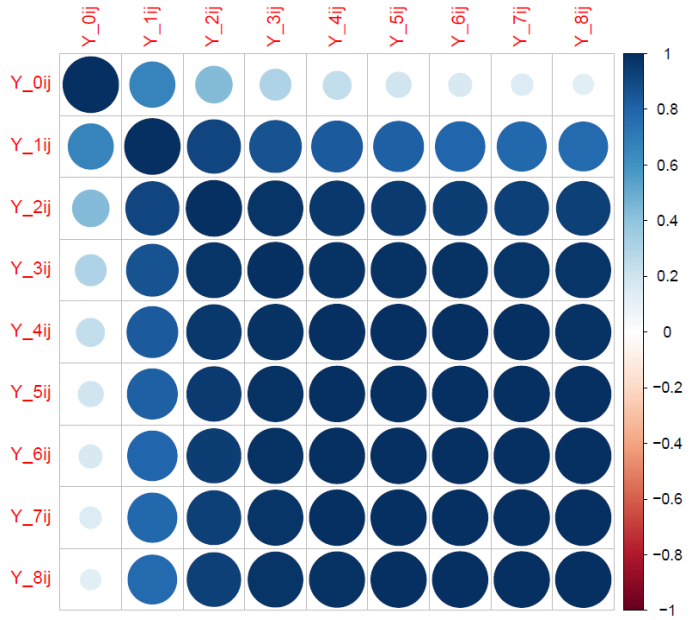
Los resultados del **modelo 8** donde la suma de las varianzas del nivel dos era un 95% de la varianza total del modelo y el 5% de varianza restante, se fijó en el error intrasujeto (estos porcentajes totales son los mismos que los fijados en el modelo 3 y 5), se presentan a continuación.

Tabla VI.24. *Matriz de correlaciones del modelo 8*

MODELO 8		
EFECTOS FIJOS		
Coef.		
β_{000}	4,6	
β_{100}	0,4	
EFECTOS ALEATORIOS		
NIVEL 3	$\sigma(u_{00j})$	$\sigma(u_{10j})$
$\sigma(u_{00j})$	0	
$\sigma(u_{10j})$		0
NIVEL 2	$\sigma(r_{0ij})$	$\sigma(r_{1ij})$
$\sigma(r_{0ij})$	**47,5%**	
$\sigma(r_{1ij})$		**47,5%**
NIVEL 1	$\sigma(e_{tij})$	
$\sigma(e_{tij})$	**5%**	
R_{aulas}	0	
$R_{estudiantes}$	0	

MATRIZ DE CORRELACIONES

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,6606**	**0,4312**	**0,3079**	**0,2412**	**0,1955**	**0,1657**	**0,1437**	**0,1272**
Y_{1ij}	0,6606	1	**0,9171**	**0,8682**	**0,8359**	**0,8105**	**0,7935**	**0,7804**	**0,7702**
Y_{2ij}	0,4312	0,9171	1	**0,9742**	**0,963**	**0,9522**	**0,944**	**0,9372**	**0,9319**
Y_{3ij}	0,3079	0,8682	0,9742	1	0,989	0,9851	0,9813	0,9778	0,9748
Y_{4ij}	0,2412	0,8359	0,963	0,989	1	0,9937	0,9923	0,9904	0,9887
Y_{5ij}	0,1955	0,8105	0,9522	0,9851	0,9937	1	0,996	0,9953	0,9945
Y_{6ij}	0,1657	0,7935	0,944	0,9813	0,9923	0,996	1	0,9973	0,9969
Y_{7ij}	0,1437	0,7804	0,9372	0,9778	0,9904	0,9953	0,9973	1	0,9979
Y_{8ij}	0,1272	0,7702	0,9319	0,9748	0,9887	0,9945	0,9969	0,9979	1



Lo primero que destaca de los resultados de este modelo, tal y como se resalta en negrita, es que se observa nuevamente una cierta tendencia a un patrón decreciente. Esta tendencia es muy marcada, sobre todo en las correlaciones de las dos primeras ocasiones de medida con el resto de ocasiones. También dicha tendencia se observa entre las correlaciones de la ocasión tres con el resto de mediciones, pero es menos marcada. A partir de esta ocasión el patrón decreciente se disipa volviéndose más constante. Así, mientras que en la segunda medición la correlación entre esta aplicación y la tres es de 0,917 y con la nueve es de 0,770, en la tercera medición, la correlación de esta ocasión con la cuatro es de 0,974 y con la

nueve es de 0,931. En el caso de la cuarta ocasión todos los valores de correlación se encuentran en torno a 0,98.

Dicho patrón decreciente entre la primera ocasión de medida y el resto de las ocasiones ya se había estudiado en algunos modelos de la situación 7. Pero por primera vez, es en este modelo 8, cuando dicho patrón se mantiene de manera tan marcada más allá de las correlaciones entre la primera ocasión de medida. Ya que generalmente, en los modelos estudiados con anterioridad predominaba un patrón constante.

El patrón decreciente o constante, que se observa en dicho modelo 8, ocurre dentro de cada ocasión de medida (por ejemplo dentro de la primera y segunda medición el patrón tiende a ser decreciente), no al comparar las correlaciones entre unas ocasiones con otras, ya que en este caso, a medida que se incrementa el número de aplicaciones aumenta el valor de la correlación, debido a al carácter acumulativo de la varianza de la pendiente, por ejemplo con cuatro mediciones la correlación es superior a la generada con tres, y con cinco mediciones es superior a la que se obtiene con cuatro, y así sucesivamente.

En el modelo 8 los coeficientes aleatorios del nivel de estudiante suman un 95% de la varianza total del modelo, mientras que el error intrasujeto es un 5%. Estos valores invertidos entre el nivel dos y la varianza del error intrasujeto se fijaron en el modelo 7, por ello no es de extrañar que los valores de correlación obtenidos en este modelo 7 sean muy inferiores a los del modelo 8, sobre todo cuando el número de mediciones es pequeño. Así, por ejemplo, mientras que en el modelo 8 la correlación entre la ocasión dos y tres es de 0,917, en el modelo 7 era de 0,079. Por lo tanto, cuando el porcentaje de la varianza del error intrasujeto es grande en comparación con el porcentaje fijado en el resto de coeficientes aleatorios, se requieren un gran número de medidas para poder predecir las puntuaciones de los estudiantes de manera fiable.

Si se comparan los resultados del modelo analizado (modelo 8), con los del modelo 3, donde la varianza de la pendiente se fijó a un 95%, los valores de correlación entre la primera ocasión y el resto de ocasiones son inferiores a las del modelo 8, ya que, en el modelo 3 todos los estudiantes partían del mismo nivel de logro. Donde sí se fijó varianza en el intercepto fue en el modelo 5, concretamente de

un 95%, por ello, en el mismo, las correlaciones entre la primera medición y el resto de las ocasiones son superiores a 0,90 y, por lo tanto, superiores a las del modelo 8. Esta intensidad de correlación se mantiene a lo largo de toda la matriz de datos, lo que supone que, dicho modelo refleja valores superiores a los del modelo 8, más allá de las correlaciones entre la primera ocasión y el resto de mediciones, a pesar de no haberse fijado varianza en la pendiente. Esto es así, porque al fijar el 95% de varianza en el intercepto, se generan unas correlaciones muy elevadas entre la primera ocasión de medida y el resto de ocasiones, que se mantienen constante a lo largo de todas las mediciones, tal y como se ha explicado en los modelos de la subsituación 4 y 5.

En cuanto al resto de correlaciones, después de la primera ocasión de medida, en el modelo 3 son ligeramente superiores a las del modelo 8, sobre todo en aquellas mediciones donde el patrón decreciente del modelo 8 es más marcado (ocasión dos). Pero en el resto de mediciones la fiabilidad obtenida en las puntuaciones de los estudiantes es muy similar en ambos modelos, a pesar de que en el modelo 8 la varianza de la pendiente es un 47,5% de la varianza total, y en el modelo 3 era un 95%. La correlación entre la ocasión dos y tres en el modelo 8 es de 0,917 y, en el modelo 3, era de 0,968, más parecidas son aún las correlaciones entre la ocasión ocho y nueve ya que en el modelo 3 era de 0,999, mientras que en el modelo 8 es de 0,997.

Estos resultados, revelan la importancia del porcentaje fijado en la varianza del error intrasujeto, ya que, en ambos modelos se obtienen unas correlaciones muy similares debido a que la varianza fijada en el error intrasujeto es la misma, aunque la varianza de la pendiente del nivel 8 sea muy inferior a la fijada en el modelo 3.

Una manera de comprobar la importancia del error intrasujeto en la fuerza de relación entre las ocasiones de medida, es comparando los resultados del modelo 8, con los del modelo 2.1, cuando el número de ocasiones no es elevado.

En este modelo 2.1 el porcentaje de varianza de la pendiente se fijó a un valor muy similar al del modelo 8, concretamente a un 35%, pero se obtienen valores de correlación mucho más bajos que los obtenidos en el modelo 8, cuando no hay tantas diferencias entre los porcentajes fijados en ambas pendientes. Así por ejemplo la correlación entre la ocasión dos y tres en el modelo 2.1 era de 0,491, mientras que en el modelo 8 es de 0,917. Esta diferencia tan marcada entre dichos valores, solo es

explicable por el porcentaje de varianza fijado en el error intrasujeto, ya que, los porcentajes de varianza fijados en la pendiente no difieren en más de trece puntos porcentuales. Sin embargo, el porcentaje de varianza fijado en el error intrasujeto en el modelo 2.1, era de un 65% y en el modelo 8 es de 5%, esta diferencia, de más de 60 puntos porcentuales entre ambos valores de la varianza, es lo que explica que las correlaciones en el modelo 2.1, sean muy inferiores a las del modelo 8, cuando el número de mediciones es bajo. Cuando el número de aplicaciones es alto, en torno a cinco, seis medidas, sí se obtienen en ambos modelos valores de correlación superiores a 0,90. Dicha explicación justificaría que los valores de correlación del modelo 3 y 8 sean tan parecidos, pero que difieran tanto del modelo 2.1, cuando se tienen pocas ocasiones de medida.

Posteriormente se muestran los resultados del **modelo 8.1,** en el cual, el porcentaje de varianza fijado en el estatus y en la pendiente del nivel de estudiante era el mismo que el fijado en el error intrasujeto.

Tabla VI.25. *Matriz de correlaciones del modelo 8.1*

MODELO 8.1		
EFECTOS FIJOS		
Coef.		
β_{000}	**4,6**	
β_{100}	**0,4**	
EFECTOS ALEATORIOS		
NIVEL 3	$\sigma(u_{00j})$	$\sigma(u_{10j})$
$\sigma(u_{00j})$	0	
$\sigma(u_{10j})$		0
NIVEL 2	$\sigma(r_{0ij})$	$\sigma(r_{1ij})$
$\sigma(r_{0ij})$	**33,3%**	
$\sigma(r_{1ij})$		**33,3%**
NIVEL 1	$\sigma(e_{tij})$	
$\sigma(e_{tij})$	**33,3%**	
R_{aulas}	0	
$R_{estudiantes}$	0	

MATRIZ DE CORRELACIONES

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,4091**	**0,2966**	**0,2175**	**0,1752**	**0,1399**	**0,119**	**0,1014**	**0,0899**
Y_{1ij}	0,4091	1	**0,7147**	**0,6985**	**0,6885**	**0,6682**	**0,659**	**0,6494**	**0,6428**
Y_{2ij}	0,2966	0,7147	1	0,8604	0,8651	0,8618	0,8589	0,855	0,8527
Y_{3ij}	0,2175	0,6985	0,8604	1	**0,9242**	**0,9274**	**0,9295**	**0,9285**	**0,9283**
Y_{4ij}	0,1752	0,6885	0,8651	0,9242	1	0,9522	0,956	0,9567	0,9569
Y_{5ij}	0,1399	0,6682	0,8618	0,9274	0,9522	1	0,9673	0,9692	0,9708
Y_{6ij}	0,119	0,659	0,8589	0,9295	0,956	0,9673	1	0,9769	0,9785
Y_{7ij}	0,1014	0,6494	0,855	0,9285	0,9567	0,9692	0,9769	1	0,9821
Y_{8ij}	0,0899	0,6428	0,8527	0,9283	0,9569	0,9708	0,9785	0,9821	1

Lo primero que se observa en los resultados de este modelo, es que, se mantiene el patrón decreciente entre la primera medición y el resto de ocasiones, así como en la ocasión dos, pero ya, dicha tendencia no es tan marcada como la del modelo 8. Además, a partir de las correlaciones de la ocasión cuatro, tal y como se ha resaltado, el patrón tiende a ser más constante o incluso creciente que decreciente.

Si los resultados de este modelo 8.1, se comparan con los del modelo 8, este último modelo muestra valores de correlación superiores en toda la matriz de datos que los del modelo 8.1, incluso existiendo un patrón decreciente, dentro de las dos primeras ocasiones de medida, más marcado en el modelo 8. Esto se debe, a que, aunque no existe mucha diferencia en los porcentajes fijados en las varianzas aleatorias del nivel 2 entre ambos modelos, el porcentaje fijado en la varianza del error intrasujeto en el modelo 8.1, es muy elevado en comparación con el que se fijó en el modelo 8. Así, por ejemplo, mientras que en el modelo 8, con dos o tres ocasiones de medida es posible predecir con una adecuada fiabilidad, las puntuaciones de los estudiantes en los siguientes cursos, pues, las correlaciones son superiores a 0,90, e el modelo 8.1 con ese mismo número de mediciones las correlaciones son más bajas, en torno a 0,85.

Finalmente, dentro de la subsituación 8, se muestran los resultados del **modelo 8.2**, donde el 50% de varianza se fijó en la varianza de la pendiente del nivel de estudiante y, el otro 50% en el estatus de ese mismo nivel.

Tabla VI.26. *Matriz de correlaciones del modelo 8.2*

MODELO 8.2		
EFECTOS FIJOS		
Coef.		
β_{000}	**4,6**	
β_{100}	**0,4**	
EFECTOS ALEATORIOS		
NIVEL 3	$\sigma(u_{00j})$	$\sigma(u_{10j})$
$\sigma(u_{00j})$	0	
$\sigma(u_{10j})$		0
NIVEL 2	$\sigma(r_{0ij})$	$\sigma(r_{1ij})$
$\sigma(r_{0ij})$	**50%**	
$\sigma(r_{1ij})$		**50%**
NIVEL 1	$\sigma(e_{tij})$	
$\sigma(e_{tij})$	0	
R_{aulas}	0	
$R_{estudiantes}$	0	

MATRIZ DE CORRELACIONES

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,7132**	**0,4579**	**0,3283**	**0,2551**	**0,2088**	**0,1772**	**0,1542**	**0,1369**
Y_{1ij}	0,7132	1	**0,9497**	**0,8962**	**0,8597**	**0,8344**	**0,8162**	**0,8026**	**0,792**
Y_{2ij}	0,4579	0,9497	1	**0,9901**	**0,9764**	**0,965**	**0,9561**	**0,949**	**0,9433**
Y_{3ij}	0,3283	0,8962	0,9901	1	0,9971	0,9923	0,9878	0,9839	0,9806
Y_{4ij}	0,2551	0,8597	0,9764	0,9971	1	0,9989	0,9968	0,9947	0,9927
Y_{5ij}	0,2088	0,8344	0,965	0,9923	0,9989	1	0,9995	0,9985	0,9973
Y_{6ij}	0,1772	0,8162	0,9561	0,9878	0,9968	0,9995	1	0,9997	0,9992
Y_{7ij}	0,1542	0,8026	0,949	0,9839	0,9947	0,9985	0,9997	1	0,9998
Y_{8ij}	0,1369	0,792	0,9433	0,9806	0,9927	0,9973	0,9992	0,9998	1

En este modelo se observa que las correlaciones entre la primera ocasión de medida con el resto de las mediciones, muestran un patrón decreciente muy marcado, que se suaviza en las siguientes aplicaciones.

Al comparar los resultados de este modelo con los del modelo 3.2, donde toda la varianza se fijaba en la pendiente del nivel de estudiante, o con los del modelo 5.2, donde todo el porcentaje de varianza se había fijado en el intercepto, los resultados son muy distintos. En el modelo 8.3 los valores de correlación entre la primera ocasión de medida y el resto de mediciones son muy bajos, aunque a partir de la primera aplicación las estimaciones son muy fiables, ya que, los valores de correlación entre las puntuaciones de los sujetos en las distintas aplicaciones se incrementan, pues, a pesar de haber cierta tendencia a un patrón decreciente, con tres mediciones las correlaciones ya son superiores a 0,90.

Tras los resultados de los modelos de la subsituación 8, se puede concluir que cuando el porcentaje fijado en la varianza del error intrasujeto, es inferior a la suma de las varianzas del nivel de estudiante o del nivel de aula, las correlaciones entre la primera ocasión y el resto de las ocasiones muestran un patrón decreciente. Dicho patrón decreciente se puede mantener más allá de la primera ocasión de medida, siendo más o menos pronunciado dependiendo del tamaño de la varianza del error intrasujeto fijado; si es muy pequeño en comparación con el resto de las varianzas como ocurría en el modelo 8 y 8.3, el patrón decreciente se puede mantener de manera pronunciada más allá de la primera ocasión. Pero si no existe tanta diferencia entre la varianza fijada en el error intrasujeto, y el resto de varianzas, como es por ejemplo el caso del modelo 8.2, dicho patrón decreciente se sigue manteniendo sobre todo en la primera ocasión de medida, pero en las mediciones posteriores no es tan pronunciado.

Seguidamente, se describen los resultados de los modelos de la **subsituación 9,** en los cuales, se fijaba varianza en el error intrasujeto y, en uno de los coeficientes aleatorios (pendiente o intercepto), del nivel de estudiante o del nivel de aula.

En los resultados del **modelo 9,** que se presentan a continuación (Tabla VI.27)**,** la varianza del error intrasujeto se fijó a un 95% de la varianza total, y cada una de las varianzas de la pendiente a un 2,5%.

Tabla VI.27. *Matriz de correlaciones del modelo 9*

MODELO 9		
EFECTOS FIJOS		
Coef.		
β_{000}	**4,6**	
β_{100}	**0,4**	
EFECTOS ALEATORIOS		
NIVEL 3	$\sigma(u_{00j})$	$\sigma(u_{10j})$
$\sigma(u_{00j})$	0	
$\sigma(u_{10j})$		**2,5%**
NIVEL 2	$\sigma(r_{0ij})$	$\sigma(r_{1ij})$
$\sigma(r_{0ij})$	0	
$\sigma(r_{1ij})$		**2,5%**
NIVEL 1	$\sigma(e_{tij})$	
$\sigma(e_{tij})$	**95%**	
R_{aulas}	0	
$R_{estudiantes}$	0	

MATRIZ DE CORRELACIONES

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	**1**	**-0,0201**	**-0,0055**	**0,0043**	**0,0006**	**0,0004**	**0,0015**	**-0,0086**	**-0,0046**
Y_{1ij}	-0,0201	1	**0,0961**	0,1238	0,1629	0,1624	0,1861	0,1913	0,2036
Y_{2ij}	-0,0055	0,0961	1	0,2322	0,2697	0,3046	0,3323	0,3463	0,3687
Y_{3ij}	0,0043	0,1238	0,2322	1	0,3788	0,4169	0,4597	0,476	0,5018
Y_{4ij}	0,0006	0,1629	0,2697	0,3788	1	0,5032	0,5485	0,5714	0,5923
Y_{5ij}	0,0004	0,1624	0,3046	0,4169	0,5032	1	0,6029	0,6268	0,6563
Y_{6ij}	0,0015	0,1861	0,3323	0,4597	0,5485	0,6029	1	0,6887	0,7128
Y_{7ij}	-0,0086	0,1913	0,3463	0,476	0,5714	0,6268	0,6887	1	**0,7414**
Y_{8ij}	-0,0046	0,2036	0,3687	0,5018	0,5923	0,6563	0,7128	0,7414	1

Cuando todos los estudiantes parten del mismo nivel de rendimiento previo y crecen a distinto ritmo, y las aulas entre si cuentan con el mismo nivel de logro previo, pero difieren en su pendiente de crecimiento, se observa que las correlaciones entre la ocasión una y el resto de ocasiones son constantes, en torno a cero, ya que, no se ha fijado varianza en el intercepto. Sin embargo, el resto de valores de correlación son superiores a los obtenidos en la primera ocasión de medida. Estos valores se van incrementando según se realizan más mediciones, aunque debido a que el porcentaje total fijado en la varianza de la pendiente, es muy pequeño (5%), en comparación con el fijado en el error intrasujeto (95%), con nueve ocasiones de medida no se pueden predecir las puntuaciones de los estudiantes con una gran precisión, ya que, la correlación entre la ocasión ocho y nueve es de 0,741.

Si se compara los resultados de este modelo 9, con el modelo 2, donde la varianza de la pendiente del nivel de estudiante se fijó a un 5% y la varianza del error intrasujeto a un 95%, se obtienen valores de correlación muy similares, ya que, los porcentajes totales de la pendiente, son los mismos en ambos modelos. Así, por ejemplo, mientras que en el modelo 2, la correlación entre la ocasión dos y tres era de 0,094, en el modelo 9 es de 0,096, la correlación entre la ocasión ocho y nueve, en el modelo 2 era de 0,739 y, en el modelo 9 es de 0,741.

En el Anexo III se muestran los resultados del **modelo 9.1,** en el cual se fijó la varianza del intercepto del nivel de aula a un 2,5% y la de la pendiente a otro 2,5%, el resto de varianza se fijó en el error intrasujeto (95%). En este modelo, como las

aulas partian de distintos niveles de logro previo, las correlaciones entre la primera ocasión de medida y el resto de mediciones son bajas (se fijó sólo un 2,5% en la varianza del estatus), pero superiores a las del modelo anterior, ya que, no sé fijó varianza en el intercepto. A partir de la segunda medida, las correlaciones son superiores en el modelo 9, ya que, se fijó un porcentaje superior en la varianza de la pendiente, que en el modelo 9.1

También en el Anexo III, se presentan los resultados del **modelo 9.2** que es similar al modelo 9, pero referido al estatus, es decir, tanto la varianza del intercepto del nivel de aula como la del nivel de estudiante se fijaron a un 2,5%, y la varianza del error intrasujeto a un 95%. Los resultados obtenidos en dicho modelo son muy parecidos a los del modelo 4, en el cual, se fijó la varianza del intercepto del nivel de aula a un 5% y, la varianza del error intrasujeto a 95%.

Tras los resultados de los modelos de la subsituación 9 se observa, nuevamente, que el nivel donde se fija varianza (nivel dos o tres), no afecta a la fiabilidad de las estimaciones; lo importante es el porcentaje de varianza total fijado en los coeficientes aleatorios. Así, por ejemplo, en un modelo donde solo se fija variabilidad en la pendiente del estudiante (modelo 2), se obtienen correlaciones similares a las de un modelo en el cual, se fijó tanto varianza en la pendiente del nivel de estudiante como del nivel de aula (modelo 9), siempre y cuando la suma de los porcentajes de las varianzas de ambas pendiente sea igual, que el porcentaje que se ha había fijado en el modelo donde solo se establecía varianza en la pendiente del nivel de estudiante, y cuando el porcentaje de varianza fijado en el error intrasujeto de ambos modelos sea el mismo.

Aunque, tanto en el modelo 2, como en el modelo 9, los resultados son muy similares, la interpretación de los mismos es muy distinta. En el modelo 2, el crecimiento académico de los estudiantes se debe exclusivamente a las características específicas de los mismos, ya que, solo se fijó varianza en la pendiente del nivel 2. En el modelo 9, el rendimiento académico de los estudiantes, no se debe solo a sus propias características, sino, también, al efecto de las aulas, ya que, se fijó varianza en la pendiente de ambos niveles. Por lo tanto, resultados similares se pueden deber a efectos distintos.

VI.1.2.2. Resultados de los modelos de la situación b

Los resultados de los modelos desde la **subsituación 10** a la **subsituación 14**, pretenden dar respuesta, dentro del primer problema de investigación, al objetivo específico 2, ***situación b***. En los modelos de esta situación, además de fijar variabilidad en el error intrasujeto y en el intercepto y la pendiente de un mismo nivel, se fijó también distintos valores de correlación entre el intercepto y la pendiente de ese mismo nivel.

Hay que recordar que todos los modelos que se desarrollan a continuación, son modelos simulados anteriormente en la *situación a*. Concretamente se trata de aquellos modelos donde se fijaba variabilidad en la pendiente y en el intercepto del mismo nivel, además de en el error intrasujeto (modelos de la subsituación 7 y 8). La diferencia de estos modelos con los de la *situación b*, es que ahora se fija un valor de correlación entre el estatus y la pendiente de uno de los niveles (estudiante o aula), que es distinto de 0.

Tal y como se explicó en el *apartado V.3.3.3. Procedimiento de generación de modelos*, dentro de la *situación b,* se fijan como valores de correlación entre el intercepto y la pendiente, valores positivos, de intensidad baja (0,2) y alta (0,9) O valores negativos (-0,2 y -0,9, modelos compensatorios).

Los resultados de los modelos en los cuales se fija una correlación positiva aparecen en una misma tabla y, del mismo modo para los modelos de correlación negativa, siempre y cuando dichos modelos cuenten con los mismos porcentajes de varianzas fijados en los mismos coeficientes aleatorios.

A continuación, dentro de la **subsituación 10**, se muestran los resultados de los modelos cuyos porcentajes de varianza fijados son los mismos que los del modelo 7, es decir, la varianza del intercepto y la pendiente del nivel de estudiante se fijaron a un 2,5%, y la varianza del error intrasujeto a un 95%. (Tabla VI.28 y Tabla VI.29).

En la siguiente tabla (Tabla VI.28) se muestra los resultados del **modelo 10** y del **modelo 10.1.** En el primer modelo la correlación entre el intercepto y la pendiente del nivel de estudiante se fijó a 0,2 y en el modelo 10.1 a 0,9.

Tabla VI.28. *Matriz de correlaciones del modelo 10 y del modelo 10.1*

MODELO 10 y 10.1		
EFECTOS FIJOS		
Coef.		
$\boldsymbol{\beta_{000}}$	**4,6**	
$\boldsymbol{\beta_{100}}$	**0,4**	
EFECTOS ALEATORIOS		
NIVEL 3	$\sigma(u_{00j})$	$\sigma(u_{10j})$
$\boldsymbol{\sigma(u_{00j})}$	0	
$\boldsymbol{\sigma(u_{10j})}$		0
NIVEL 2	$\sigma(r_{0ij})$	$\sigma(r_{1ij})$
$\boldsymbol{\sigma(r_{0ij})}$	**2,5%**	
$\boldsymbol{\sigma(r_{1ij})}$		**2,5%**
NIVEL 1	$\sigma(e_{tij})$	
$\boldsymbol{\sigma(e_{tij})}$	**95%**	
$\boldsymbol{R_{aulas}}$	0	
$\boldsymbol{R_{estudiantes}}$	**0,2/0,9**	

MATRIZ DE CORRELACIONES

MODELO 10

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,0347**	**0,0463**	**0,0426**	**0,0302**	**0,0508**	**0,0379**	**0,0323**	**0,0331**
Y_{1ij}	0,0347	1	**0,1008**	0,1206	0,1299	0,154	0,1614	0,1567	0,1816
Y_{2ij}	0,0463	0,1008	1	0,1662	0,2005	0,2155	0,2437	0,2485	0,2697
Y_{3ij}	0,0426	0,1206	0,1662	1	0,2791	0,3146	0,3413	0,3667	0,3901
Y_{4ij}	0,0302	0,1299	0,2005	0,2791	1	0,3794	0,4006	0,4302	0,4533
Y_{5ij}	0,0508	0,154	0,2155	0,3146	0,3794	1	0,4783	0,4874	0,5103
Y_{6ij}	0,0379	0,1614	0,2437	0,3413	0,4006	0,4783	1	0,5418	0,5736
Y_{7ij}	0,0323	0,1567	0,2485	0,3667	0,4302	0,4874	0,5418	1	**0,6103**
Y_{8ij}	0,0331	0,1816	0,2697	0,3901	0,4533	0,5103	0,5736	0,6103	1

MODELO 10.1

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,0479**	**0,0748**	**0,0776**	**0,0949**	**0,1071**	**0,1146**	**0,1224**	**0,1224**
Y_{1ij}	0,0479	1	**0,137**	0,1642	0,1957	0,206	0,2363	0,2487	0,2499
Y_{2ij}	0,0748	0,137	1	0,233	0,2726	0,3076	0,3285	0,3454	0,3482
Y_{3ij}	0,0776	0,1642	0,233	1	0,3418	0,3805	0,4121	0,429	0,4484
Y_{4ij}	0,0949	0,1957	0,2726	0,3418	1	0,4343	0,4714	0,5	0,5152
Y_{5ij}	0,1071	0,206	0,3076	0,3805	0,4343	1	0,51	0,5488	0,5676
Y_{6ij}	0,1146	0,2363	0,3285	0,4121	0,4714	0,51	1	0,5956	0,6172
Y_{7ij}	0,1224	0,2487	0,3454	0,429	0,5	0,5488	0,5956	1	**0,6432**
Y_{8ij}	0,1224	0,2499	0,3482	0,4484	0,5152	0,5676	0,6172	0,6432	1

En estos modelos donde se fijó una correlación positiva entre el estatus y la pendiente en el nivel de estudiante, se considera que, en general, los estudiantes que parten de niveles de rendimiento alto su pendiente crece mucho a lo largo del tiempo, en comparación con las pendientes de crecimiento de los estudiantes que parten de un nivel de logro bajo, ya que, en promedio, su pendiente de crecimiento no se incrementa mucho. Dicha tendencia es mas marcada cuando se fija una correlación entre ambos coeficientes de 0,9 que de 0,2.

Si se comparan los resultados del **modelo 10**, donde la correlación entre el intercepto y la pendiente del nivel de aula se fijó a 0,2, con los resultados del modelo

7 donde dicha correlación se fijó a 0, se observa que los valores de correlación del modelo 10 son ligeramente superiores a los del modelo 7. Por ejemplo, en el modelo 7, la correlación entre la ocasión dos y tres era de 0,079 y en el modelo 10 es de 0,100, o por ejemplo en este último modelo la correlación entre la ocasión ocho y nueve es de 0,610 y en el modelo 7 era de 0,589.

Al comparar los resultados del **modelo 10.1,** donde la fuerza de asociación entre el intercepto y la pendiente del nivel de estudiantes es alta (0,9), con los resultados del modelo 10, donde dicho valor de correlación es bajo, se esperarían correlaciones muy superiores en el modelo 10.1, respecto al modelo 10, pero, tal y como se puede observar, en la Tabla VI.28, ambos modelos con el mismo número de ocasiones de medida, cuentan con una fiabilidad muy parecida. Por ejemplo, mientras que en el modelo 10, la correlación entre la ocasión dos y tres es de 0,100, en el modelo 10.1, es de 0,137. Que no exista mucha diferencia en la precisión con la que se estiman las puntuaciones de los estudiantes entre ambos modelos, puede ser debido a que, el porcentaje de varianza fijado en el error intrasujeto es muy alto.

Donde más difieren, el modelo 10 y el modelo 10.1, tal y como se ha resaltado en la tabla, es en el patrón de correlaciones entre la primera ocasión de medida y el resto de ocasiones. Mientras que en el modelo 10 se observa un patrón constante (recuérdese que en el modelo 7 tendia a decreciente), en el modelo 10.1, por primera vez se observa un patrón con una cierta tendencia creciente, lo que hace que en esta primera ocasión los valores de correlación entre ambos modelos difieran más que en el resto de aplicaciones, así, por ejemplo, en el modelo 10 la correlación entre la primera ocasión y la nueve es de 0,033 y en el modelo 10.1 es de 0,122.

En la tabla siguiente, se muestra los resultados de los modelos en los cuales se fijó una correlación negativa entre el intercepto y la pendiente. En el modelo 10.2, la correlación entre el intercepto y la pendiente del nivel de estudiante se fijó a -0,2, y en el modelo 10.3, la correlación entre el estatus y la pendiente se fijó a -0,9. Estos modelos se pueden considerar compensatorios, ya que, al fijar una correlación negativa entre el estatus y la pendiente del nivel de estudiante, se representa una situación en la cual, en general, los estudiantes que parten de niveles de logro bajo incrementan mucho su rendimiento a lo largo del tiempo, mientras que los que parten de niveles de rendimiento alto mantienen sus niveles de logro en las diferentes ocasiones de medida, o bien este rendimiento no crece mucho. De tal manera que las

caracteristicas del propio sistema educativo, hacen que a medida que se realizan mas mediciones, se compensen las diferencias entre los estudiantes que parten de rendimiento alto y los que parten de rendimiento bajo, ya que, se tiende a igualar los niveles de logro promedio de ambos grupos de estudiantes.

Esta situación compensatoria es sobre todo más marcada, cuando la correlación negativa entre el intercepto y la pendiente del nivel de estudiante tiene una fuerza de relación alta (modelo 10.3) que cuando es baja (modelo 10.3). Además, en los mismos es posible que se de el efecto techo, ya que, el rendimiento de los estudiantes que parten de niveles de logro alto no puede crecer más.

Tabla VI.29. *Matriz de correlaciones del modelo 10.2 y del modelo 10.3*

MODELO 10.2 y 10.3		
EFECTOS FIJOS		**MATRIZ DE CORRELACIONES**
Coef.		**MODELO 10.2**
β_{000}	**4,6**	
β_{100}	**0,4**	
EFECTOS ALEATORIOS		
NIVEL 3	$\sigma(u_{00j})$ $\sigma(u_{10j})$	
$\sigma(u_{00j})$	0	
$\sigma(u_{10j})$	0	
NIVEL 2	$\sigma(r_{0ij})$ $\sigma(r_{1ij})$	
$\sigma(r_{0ij})$	**2,5%**	
$\sigma(r_{1ij})$	**2,5%**	
NIVEL 1	$\sigma(e_{tij})$	
$\sigma(e_{tij})$	**95%**	
R_{aulas}	0	
$R_{estudiantes}$	**-0,2/-0,9**	

MODELO 10.2

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	0,013	0,0117	0,0179	-0,0034	0,0112	-0,0006	-0,0075	-0,0106
Y_{1ij}	0,013	1	**0,0516**	0,0728	0,0777	0,0796	0,0861	0,1024	0,1141
Y_{2ij}	0,0117	0,0516	1	0,124	0,1618	0,1749	0,2048	0,2166	0,2261
Y_{3ij}	0,0179	0,0728	0,124	1	0,2278	0,2555	0,284	0,3062	0,3179
Y_{4ij}	-0,0034	0,0777	0,1618	0,2278	1	0,3215	0,3578	0,3941	0,4058
Y_{5ij}	0,0112	0,0796	0,1749	0,2555	0,3215	1	0,4135	0,4596	0,4828
Y_{6ij}	-0,0006	0,0861	0,2048	0,284	0,3578	0,4135	1	0,5104	0,5322
Y_{7ij}	-0,0075	0,1024	0,2166	0,3062	0,3941	0,4596	0,5104	1	**0,5849**
Y_{8ij}	-0,0106	0,1141	0,2261	0,3179	0,4058	0,4828	0,5322	0,5849	1

MODELO 10.3

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	-0,0156	-0,0217	-0,033	-0,0543	-0,0707	-0,0824	-0,1058	-0,1091
Y_{1ij}	-0,0156	1	**0,0135**	0,0079	0,0287	0,0078	0,0222	0,0174	0,0254
Y_{2ij}	-0,0217	0,0135	1	0,0581	0,0701	0,0916	0,1096	0,1161	0,1388
Y_{3ij}	-0,033	0,0079	0,0581	1	0,1435	0,1696	0,2099	0,2202	0,2501
Y_{4ij}	-0,0543	0,0287	0,0701	0,1435	1	0,2459	0,2917	0,3134	0,337
Y_{5ij}	-0,0707	0,0078	0,0916	0,1696	0,2459	1	0,348	0,3728	0,4127
Y_{6ij}	-0,0824	0,0222	0,1096	0,2099	0,2917	0,348	1	0,4528	0,4857
Y_{7ij}	-0,1058	0,0174	0,1161	0,2202	0,3134	0,3728	0,4528	1	**0,5223**
Y_{8ij}	-0,1091	0,0254	0,1388	0,2501	0,337	0,4127	0,4857	0,5223	1

El **modelo 10.2**, donde la correlación entre el intercepto y la pendiente se fijó a -0,2, muestra valores de correlación algo inferiores a los del modelo 10, y, por supuesto, al modelo 10.1. Así por ejemplo mientras que en el modelo 10 la correlación entre la ocasión ocho y nueve era de 0,610, en el modelo 10.2 es de 0,584. Por lo tanto, con el mismo número de mediciones, y con los mismos porcentajes de varianza fijados en ambos modelos, el modelo compensatorio es ligeramente menos preciso, que el modelo 10.

El **modelo 10.3**, muestra los valores de correlación más bajos de los cuatro modelos de esta subsituación, debido a que se fijó una correlación entre el estatus y la pendiente muy alta con direccionalidad negativa (-0,9). Aunque cuando el número de ocasiones es elevado, los resultados de dicho modelo no difieren mucho de los del resto de modelos de esta misma subsituación, ni si quiera del modelo 10.1, donde la correlación entre el intercepto y la pendiente se fijó a 0,9, pues, en dicho modelo con nueve ocasiones de medida la correlación era de 0,643 y en el modelo 10.4 es de 0,522.

Bajo estas condiciones donde la varianza fijada en el error intrasujeto es muy alta, en comparación con el resto de coeficientes aleatorios, los valores de correlación fijados entre el intercepto y la pendiente, cuando el número de mediciones es elevado no afecta de manera muy marcada a la fiabilidad de las estimaciones. Cuando el número de ocasiones son tres o cuatro, las diferencias son algo más patentes, obteniéndose los valores de correlación más altos y, por tanto, más fiables, en el modelo donde la correlación es alta positiva (modelo 10.1) y, los valores más bajos en el modelo 10.3, donde la correlación entre el intercepto y la pendiente se fijó a -0,9.

En cuanto al patrón de correlación entre la primera ocasión de medida y el resto de mediciones, hay una tendencia hacia un patrón nulo, constante, en torno a 0, como en el modelo 7. Sólo cuando se fija una correlación entre el intercepto y la pendiente, alta y positiva, como es la del modelo 10.1, dicho patrón tiene una ligera tendencia hacia el patrón creciente, lo que supone que a medida que se realizan más aplicaciones, el ERM se va mitigando, pudiendo llegar a desaparecer si la correlación es perfecta.

A continuación, se muestran los resultados de los modelos de la **subsituación 11** (Tabla VI.30 y Tabla VI.31) donde el porcentaje de varianza fijado en el intercepto era un 40% de la varianza total del modelo, y el de la pendiente un 10%, el resto de varianza se fijó en el error intrasujeto (50%). Por lo tanto, son los mismos porcentajes de varianza que los fijados en el modelo 7.2.

Los resultados de los **modelos 11 y 11.1**, se presentan a continuación, en ellos la correlación entre el estatus y la pendiente del nivel dos, se ha fijado a 0,2 y 0,9, respectivamente.

Tabla VI.30. *Matriz de correlaciones del modelo 11 y del modelo 11.1*

MODELO 11 y 11.1			
EFECTOS FIJOS			**MATRIZ DE CORRELACIONES**
Coef.			**MODELO 11**
$\boldsymbol{\beta_{000}}$	**4,6**		
$\boldsymbol{\beta_{100}}$	**0,4**		
EFECTOS ALEATORIOS			
NIVEL 3	$\sigma(u_{00j})$	$\sigma(u_{10j})$	
$\boldsymbol{\sigma(u_{00j})}$	0		
$\boldsymbol{\sigma(u_{10j})}$		0	
NIVEL 2	$\sigma(r_{0ij})$	$\sigma(r_{1ij})$	
$\boldsymbol{\sigma(r_{0ij})}$	**40%**		
$\boldsymbol{\sigma(r_{1ij})}$		**10%**	
NIVEL 1	$\sigma(e_{tij})$		
$\boldsymbol{\sigma(e_{tij})}$	**50%**		
$\boldsymbol{R_{aulas}}$	0		
$\boldsymbol{R_{estudiantes}}$	**0,2/0,9**		

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,4527**	**0,4256**	**0,3857**	**0,3507**	**0,3289**	**0,2965**	**0,2824**	**0,2638**
Y_{1ij}	0,4527	1	**0,5721**	0,5712	0,5645	0,5548	0,5377	0,528	0,5207
Y_{2ij}	0,4256	0,5721	1	0,6897	0,7104	0,7094	0,7077	0,7055	0,7015
Y_{3ij}	0,3857	0,5712	0,6897	1	0,7825	0,794	0,7969	0,8024	0,8016
Y_{4ij}	0,3507	0,5645	0,7104	0,7825	1	0,8456	0,8513	0,8594	0,8608
Y_{5ij}	0,3289	0,5548	0,7094	0,794	0,8456	1	0,8839	0,8888	0,8918
Y_{6ij}	0,2965	0,5377	0,7077	0,7969	0,8513	0,8839	1	0,9083	0,9145
Y_{7ij}	0,2824	0,528	0,7055	0,8024	0,8594	0,8888	0,9083	1	**0,9271**
Y_{8ij}	0,2638	0,5207	0,7015	0,8016	0,8608	0,8918	0,9145	0,9271	1

MODELO 11.1

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	**1**	**0,5322**	**0,5735**	**0,5862**	**0,5982**	**0,6082**	**0,6101**	**0,6142**	**0,6132**
Y_{1ij}	0,5322	1	**0,6962**	0,7205	0,7387	0,7484	0,757	0,762	0,7618
Y_{2ij}	0,5735	0,6962	1	0,79	0,8115	0,8275	0,8339	0,8399	0,8405
Y_{3ij}	0,5862	0,7205	0,79	1	0,8486	0,8648	0,8732	0,8789	0,8829
Y_{4ij}	0,5982	0,7387	0,8115	0,8486	1	0,8895	0,8991	0,9068	0,9101
Y_{5ij}	0,6082	0,7484	0,8275	0,8648	0,8895	1	0,9145	0,9236	0,9275
Y_{6ij}	0,6101	0,757	0,8339	0,8732	0,8991	0,9145	1	0,9344	0,9387
Y_{7ij}	0,6142	0,762	0,8399	0,8789	0,9068	0,9236	0,9344	1	**0,9454**
Y_{8ij}	0,6132	0,7618	0,8405	0,8829	0,9101	0,9275	0,9387	0,9454	1

En los modelos de la tabla anterior VI.30, se observa que cuando se fija una correlación baja positiva entre el intercepto y la pendiente del nivel de estudiante (modelo 11), el patrón de correlación entre la primera ocasión y el resto de ocasiones es el mismo que cuando no se fijá correlación (modelo 7.2), es decir, un patrón con tendencia decreciente. Sin embargo, cuando se fija una correlación alta positiva (modelo 11.1) el patrón entre la primera ocasión de medida y el resto de ocasiones tiende a ser creciente, mostrando valores de correlación superiores a los del modelo 7.2 y modelo 11.

Si se compara los resultados del **modelo 11** y del **modelo 11.1**, no existe una diferencia muy marcada entre las correlaciones obtenidas en ambos modelos, a excepción de las correlaciones entre la primera ocasión y el resto de ocasiones, que mientras que en el modelo 11 decrecen, en el modelo 11.1 crecen. Así, por ejemplo, en el modelo 11 la correlación entre la primera aplicación y la nueve es de 0,263, y en el modelo 11.1 es de 0,613. Sin embargo, con tres ocasiones de medida, en el modelo 11 la correlación es en torno a 0,70 y en el modelo 11.1, es cercana a 0,80, y ya con nueve mediciones las correlaciones en ambos modelos son superiores a 0,90.

Los resultados del **modelo 11.2,** donde se fijan los mismos porcentajes de varianza que en el modelo 11.1, pero referido al nivel de aula en vez de al nivel de estudiante, y, por lo tanto, el valor de la correlación (0,9) se fija en el nivel tres, se muestran en el Anexo III, ya que, los resultados son muy similares a los comentados para el modelo 11.1, aunque la interpretación de los mismos es muy distinta. Pues mientras que en el modelo 11.2, el rendimiento de los estudiantes se debe a las características de las aulas, en el modelo 11.1 es consecuencia de las características sistemáticas de los estudiantes.

En la tabla siguiente (VI.31) aparecen los resultados de los modelos de la subsituación 11 en los cuales se fijó la correlación entre el intercepto y la pendientedel nivel de estudiante a -0,2 y -0,9, respectivamente.

Tabla VI.31. *Matriz de correlaciones del modelo 11.3 y del modelo 11.5*

MODELO 11.3 y 11.5

EFECTOS FIJOS			MATRIZ DE CORRELACIONES
Coef.			
β_{000}	**4,6**		
β_{100}	**0,4**		
EFECTOS ALEATORIOS			
NIVEL 3	$\sigma(u_{00j})$	$\sigma(u_{10j})$	
$\sigma(u_{00j})$	0		
$\sigma(u_{10j})$		0	
NIVEL 2	$\sigma(r_{0ij})$	$\sigma(r_{1ij})$	
$\sigma(r_{0ij})$	**40%**		
$\sigma(r_{1ij})$		**10%**	
NIVEL 1	$\sigma(e_{tij})$		
$\sigma(e_{tij})$	**50%**		
R_{aulas}	0		
$R_{estudiantes}$	**-0,2/-0,9**		

MODELO 11.3

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	**1**	**0,3943**	**0,3125**	**0,2469**	**0,173**	**0,1303**	**0,0899**	**0,0601**	**0,0386**
Y_{1ij}	0,3943	1	0,4628	0,4531	0,42	0,3893	0,3663	0,3512	0,3384
Y_{2ij}	0,3125	0,4628	1	0,5977	0,6093	0,6021	0,599	0,5901	0,5818
Y_{3ij}	0,2469	0,4531	0,5977	1	0,7173	0,7287	0,7332	0,7332	0,7297
Y_{4ij}	0,173	0,42	0,6093	0,7173	1	0,7954	0,8086	0,8169	0,8166
Y_{5ij}	0,1303	0,3893	0,6021	0,7287	0,7954	1	0,8495	0,8632	0,868
Y_{6ij}	0,0899	0,3663	0,599	0,7332	0,8086	0,8495	1	0,8892	0,8953
Y_{7ij}	0,0601	0,3512	0,5901	0,7332	0,8169	0,8632	0,8892	1	0,9148
Y_{8ij}	0,0386	0,3384	0,5818	0,7297	0,8166	0,868	0,8953	0,9148	1

MODELO 11.5

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,2779**	**0,0599**	**-0,1636**	**-0,3171**	**-0,41**	**-0,4634**	**-0,5016**	**-0,5212**
Y_{1ij}	0,2779	1	0,1118	-0,0241	-0,098	-0,1728	-0,201	-0,2265	-0,2392
Y_{2ij}	0,0599	0,1118	1	**0,1576**	**0,1542**	**0,1444**	**0,1382**	**0,1286**	**0,1289**
Y_{3ij}	-0,1636	-0,0241	0,1576	1	0,389	0,4192	0,4455	0,4487	0,4584
Y_{4ij}	-0,3171	-0,098	0,1542	0,389	1	0,5993	0,6396	0,6553	0,6656
Y_{5ij}	-0,41	-0,1728	0,1444	0,4192	0,5993	1	0,7308	0,7524	0,7713
Y_{6ij}	-0,4634	-0,201	0,1382	0,4455	0,6396	0,7308	1	0,8177	0,8347
Y_{7ij}	-0,5016	-0,2265	0,1286	0,4487	0,6553	0,7524	0,8177	1	**0,8646**
Y_{8ij}	-0,5212	-0,2392	0,1289	0,4584	0,6656	0,7713	0,8347	0,8646	1

En los **modelos 11.3 y 11.5,** hay una tendencia hacia un patrón decreciente dentro de la primera y segunda ocasión de medida. Es más, en el modelo 11.5, donde la correlación entre el intercepto y la pendiente es de -0,9, dicha tendencia decreciente se extiende ligeramente hasta dentro de la tercera medición.

El **modelo 11.3**, presenta valores de correlación ligeramente inferiores a los del modelo 11, aunque dicha diferencia no es muy marcada, sobre todo con nueve ocasiones de medida, ya que, en ambos modelos las estimaciones de las puntuaciones verdaderas de los estudiantes son muy precisas (valores de correlación superiores a 0,90). Dicho modelo 11.3, difiere algo mas del modelo 11.1, ya que, mientras en este

último modelo con cinco o seis medidas ya se obtiene mucha información sobre los niveles de logro de los estudiantes, en el modelo 11.3 se requiren de más aplicaciones.

Los resultados de un modelo similar al 11.3, se presentan en el Anexo III. En este **modelo 11.4**, se fijaron los mismos porcentajes de varianza en el nivel de aula que los fijados en el modelo 11.3 en el nivel de estudiante. Dicho modelo se presenta en el anexo porque los resultados son muy similares a los obtenidos en el modelo 11.3.

Finalmente, en cuanto al **modelo 11.5**, es el modelo que muestra los valores de correlación más bajos, ya que, en el mismo, con 9 mediciones las correlaciones son inferiores a 0,90. Además, con tres ocasiones de medida, que son el número mínimo de mediciones que se requieren en las evaluaciones longitudinales basadas en modelos de crecimiento, se obtendrían estimaciones muy poco precisas sobre el rendimiento real de los estudiantes, puesto que las correlaciones son muy bajas (0,15).

A la vista de estos resultados, se puede concluir que la fiabilidad de las estimaciones de las puntuaciones de los estudiantes, se ve afectada por la diferencia de porcentajes fijados en los diversos coeficientes aleatorios, pero también por la intensidad y la direccionalidad de los valores de correlación fijados entre el intercepto y la pendiente, aunque no de manera muy marcada.

El modelo donde la fuerza de asociación entre el intercepto y la pendiente es alta y de direccionalidad negativa, es el que presenta valores de correlación más bajos, por el contrario, el modelo donde se fija una correlación entre el intercepto y la pendiente alta y de símbolo positivo, es el modelo que muestra más precisión en las estimaciones de las puntuaciones de los estudiantes.

En los resultados de los modelos presentados hasta el momento dentro de la *situación b*, el porcentaje de varianza fijado en el error intrasujeto, era superior o igual a la suma de los coeficientes aleatorios del nivel dos o tres. A continuación, se muestran los resultados de los modelos, donde la varianza fijada en el error intrasujeto, es inferior a la del resto de coeficientes en los que se fija variabilidad. Estos modelos también han sido simulados anteriormente, concretamente en la subsituación 8, pero en la misma no se fijaba una correlación entre el intercepto y la pendiente distinta de cero.

En los resultados de los modelos de la **subsituación 12,** que se presentan a continuación (Tabla VI.32 y Tabla VI.33), los porcentajes fijados en la varianza del error intrasujeto y en el nivel dos, son los contrarios a los de la subsituación 10. Es decir, en estos modelos el 95% de la varianza se estableció en el nivel de estudiante, fijandose cada coeficiente aleatorio a un 47,5%, y la varianza del error intrasujeto se fijó a un 5%, (mismos porcentajes que en el modelo 8).

En la tabla siguiente VI.32, se muestra los resultados del **modelo 12 y 12.1**, donde la correlación entre el intercepto y la pendiente del nivel dos se fijó a 0,2 y a 0,9, respectivamente.

Tabla VI.32. *Matriz de correlaciones del modelo 12 y del modelo 12.1*

MODELO 12 y 12.1

EFECTOS FIJOS			MATRIZ DE CORRELACIONES
Coef.			
β_{000}	**4,6**		
β_{100}	**0,4**		
EFECTOS ALEATORIOS			
NIVEL 3	$\sigma(u_{00j})$	$\sigma(u_{10j})$	
$\sigma(u_{00j})$	0		
$\sigma(u_{10j})$		0	
NIVEL 2	$\sigma(r_{0ij})$	$\sigma(r_{1ij})$	
$\sigma(r_{0ij})$	**47,5%**		
$\sigma(r_{1ij})$		**47,5%**	
NIVEL 1	$\sigma(e_{tij})$		
$\sigma(e_{tij})$	**5%**		
R_{aulas}	0		
$R_{estudiantes}$	**0,2/0,9**		

MODELO 12

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,7279**	**0,5612**	**0,4653**	**0,4086**	**0,3719**	**0,3462**	**0,3259**	**0,3116**
Y_{1ij}	0,7279	1	**0,9381**	**0,9042**	**0,8789**	**0,8606**	**0,8471**	**0,8359**	**0,828**
Y_{2ij}	0,5612	0,9381	1	0,9787	0,9701	0,9618	0,9553	0,9494	0,9451
Y_{3ij}	0,4653	0,9042	0,9787	1	0,9903	0,9874	0,9842	0,9812	0,9787
Y_{4ij}	0,4086	0,8789	0,9701	0,9903	1	0,9945	0,9932	0,9917	0,9902
Y_{5ij}	0,3719	0,8606	0,9618	0,9874	0,9945	1	0,9965	0,9958	0,995
Y_{6ij}	0,3462	0,8471	0,9553	0,9842	0,9932	0,9965	1	0,9974	0,9971
Y_{7ij}	0,3259	0,8359	0,9494	0,9812	0,9917	0,9958	0,9974	1	0,9981
Y_{8ij}	0,3116	0,828	0,9451	0,9787	0,9902	0,995	0,9971	0,9981	1

MODELO 12.1

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,9164**	**0,9051**	**0,8957**	**0,89**	**0,885**	**0,8819**	**0,879**	**0,8769**
Y_{1ij}	0,9164	1	0,9785	0,9774	0,9762	0,9744	0,9735	0,9723	0,9714
Y_{2ij}	0,9051	0,9785	1	0,99	0,9903	0,9899	0,9895	0,9889	0,9885
Y_{3ij}	0,8957	0,9774	0,99	1	0,9943	0,9945	0,9945	0,9942	0,9941
Y_{4ij}	0,89	0,9762	0,9903	0,9943	1	0,9962	0,9964	0,9965	0,9964
Y_{5ij}	0,885	0,9744	0,9899	0,9945	0,9962	1	0,9974	0,9975	0,9976
Y_{6ij}	0,8819	0,9735	0,9895	0,9945	0,9964	0,9974	1	0,9981	0,9982
Y_{7ij}	0,879	0,9723	0,9889	0,9942	0,9965	0,9975	0,9981	1	0,9985
Y_{8ij}	0,8769	0,9714	0,9885	0,9941	0,9964	0,9976	0,9982	0,9985	1

Si se observa los valores de correlación del **modelo 12,** y se comparan con el modelo 8, donde la correlación entre el intercepto y la pendiente del nivel de estudiante se fijó a 0; en el modelo 12, se obtienen valores de correlación ligeramente más elevados, sobre todo dentro de la primera ocasión de medida, ya que, por ejemplo, en este modelo la correlación entre la ocasión una y nueve, es de 0,312, y en el modelo 8 era de 0,127; pero entre el resto de medicioness la diferencia no es tan marcada pues, como en ambos modelos la intensidad de relación entre el intercepto y la pendiente no es muy alta, se obtienen correlaciones similares. Así, por ejemplo, mientras que en el modelo 8 la correlación entre la ocasión dos y tres era de 0,917, en el modelo 13 es de 0,938. Además, en el modelo 12, tal y como ocurria con el modelo 8, se mantiene el patrón decreciente en la mayoría de las ocasiones de medida, siendo, en ambos modelos, dicho patrón especialmente marcado dentro de las dos primeras ocasiones de medida.

En el modelo 12, como se ha comentado, se invierte los porcentajes totales de varianza fijados en el error intrasujeto y en el nivel de estudiantes del modelo 10, por ello, los resultados al comparar ambos modelos son muy distintos. Si se recuerda, en el modelo 10, ni con nueve ocasiones se obtenían correlaciones superiores a 0,65. Sin embargo, en el modelo 12, con tres mediciones se puede predecir el rendimiento de los estudiantes en los cursos siguientes, de manera muy fiable, ya que, las correlaciones son superiores a 0,90.

En el **modelo 12.1**, a excepción de los valores de correlación entre la primera ocasión de medida y el resto de mediciones, los valores obtenidos son muy similares a los del modelo 12, debido a que en ambos modelos se fija un porcentaje de varianza en el error intrasujeto muy pequeño en comparación con el resto de coeficientes aleatorios. Como la varianza del error intrasujeto es muy pequeña en comparación con la varianza de la pendiente, con pocas ocasiones de medida se obtienen valores de correlación altos, lo que hace que dichos valores no se puedan incrementar mucho, aunque la correlación entre el intercepto y la pendiente sea muy elevada.

Finalmente hay que destacar de dicho modelo 12.1, el patrón de correlación obtenido entre la primera ocasión de medida y el resto. Hasta el momento, siempre que se fijaba un valor de correlación entre el intercepto y la pendiente de 0,9, el patrón de correlación que se generaba entre la primera medida y el resto de ocasiones era creciente. Sin embargo, por primera vez, se encuentra un patrón con

una cierta tendencia decreciente, que puede ser consecuencia de la combinación de fijar una correlación entre el intercepto y la pendiente alta, junto con un porcentaje de varianza en el intercepto muy elevado en comparación con la varianza fijada en el error intrasujeto. Cuando se dan estas condiciones, parece que entre la primera ocasión de medida y el resto de ocasiones se genera un patrón decreciente. Para demostrar esto, se desarrolló un modelo donde se fija la varianza del intercepto del nivel de estudiante a un 85% de la varianza total del modelo, la varianza de la pendiente de ese mismo nivel a un 10% y el error intrasujeto a un 5%, con un valor de correlación entre el intercepto y la pendiente de 0,9. Como se puede observar en el Anexo III, en el cual se presentan los resultados del modelo 12.1.1, el patrón entre la primera medida y el resto de ocasiones, muestra una tendencia decreciente, debido a la combinación de los factores mencionados anteriormente.

Si el porcentaje de varianza fijado en el estatus es menor, a la varianza fijada en el error intrasujeto, con una correlación entre el intercepto y la pendiente de 0,9, el patrón que se obtendría dentro de la primera ocasión de medida sería creciente, tal y como se observó en los modelos 10.1 y 11.1

Finalmente, dentro de la subsituación 12, en la siguiente tabla VI.33, se muestra los resultados del **modelo 12.2 y 12.3**, donde las correlaciones entre el intercepto y la pendiente se fijaron a -0, 2 y -0,9, respectivamente.

Tabla VI.33. *Matriz de correlaciones del modelo 12.2 y del modelo 12.3*

MODELO 12.2 y 12.3			
EFECTOS FIJOS			**MATRIZ DE CORRELACIONES**
Coef.			**MODELO 12.2**
β_{000}	**4,6**		
β_{100}	**0,4**		
EFECTOS ALEATORIOS			
NIVEL 3	$\sigma(u_{00j})$	$\sigma(u_{10j})$	
$\sigma(u_{00j})$	0		
$\sigma(u_{10j})$		0	
NIVEL 2	$\sigma(r_{0ij})$	$\sigma(r_{1ij})$	
$\sigma(r_{0ij})$	**47,5%**		
$\sigma(r_{1ij})$		**47,5%**	
NIVEL 1	$\sigma(e_{tij})$		
$\sigma(e_{tij})$	**5%**		
R_{aulas}	0		
$R_{estudiantes}$	**-0,2/-0,9**		

MODELO 12.2

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,5899**	**0,2826**	**0,1391**	**0,0569**	**0,009**	**-0,024**	**-0,0464**	**-0,0639**
Y_{1ij}	0,5899	1	**0,8848**	**0,8219**	**0,7771**	**0,7476**	**0,7265**	**0,7118**	**0,7**
Y_{2ij}	0,2826	0,8848	1	0,9692	0,9548	0,9423	0,9326	0,9252	0,9189
Y_{3ij}	0,1391	0,8219	0,9692	1	0,9868	0,9823	0,9776	0,9738	0,9703
Y_{4ij}	0,0569	0,7771	0,9548	0,9868	1	0,993	0,9914	0,9895	0,9876
Y_{5ij}	0,009	0,7476	0,9423	0,9823	0,993	1	0,9956	0,995	0,994
Y_{6ij}	-0,024	0,7265	0,9326	0,9776	0,9914	0,9956	1	0,9971	0,9967
Y_{7ij}	-0,0464	0,7118	0,9252	0,9738	0,9895	0,995	0,9971	1	0,9979
Y_{8ij}	-0,0639	0,7	0,9189	0,9703	0,9876	0,994	0,9967	0,9979	1

MODELO 12.3

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,1627**	**-0,6153**	**-0,743**	**-0,782**	**-0,8025**	**-0,8136**	**-0,8211**	**-0,8261**
Y_{1ij}	0,1627	1	**0,4579**	**0,3452**	**0,3028**	**0,2745**	**0,2597**	**0,2489**	**0,2413**
Y_{2ij}	-0,6153	0,4579	1	0,939	0,9322	0,9255	0,9211	0,9175	0,9151
Y_{3ij}	-0,743	0,3452	0,939	1	0,9817	0,9808	0,9799	0,9787	0,9778
Y_{4ij}	-0,782	0,3028	0,9322	0,9817	1	0,991	0,9912	0,9909	0,9905
Y_{5ij}	-0,8025	0,2745	0,9255	0,9808	0,991	1	0,9947	0,9948	0,9948
Y_{6ij}	-0,8136	0,2597	0,9211	0,9799	0,9912	0,9947	1	0,9965	0,9967
Y_{7ij}	-0,8211	0,2489	0,9175	0,9787	0,9909	0,9948	0,9965	1	0,9975
Y_{8ij}	-0,8261	0,2413	0,9151	0,9778	0,9905	0,9948	0,9967	0,9975	1

En los **modelos 12.2 y 12.3**, se observa que los valores obtenidos en estas matrices de correlación, son inferiores a los que se obtienen cuando se fija una correlación entre el intercepto y la pendiente positiva.

Si se compara el modelo 12.3 con el modelo 12.1, cuando el número de ocasiones de medida es pequeño, el modelo 12.3 muestra estimaciones mucho más precisas en las puntuaciones del rendimiento de los estudiantes que el modelo 12.1. Ya que, mientras en este último modelo la correlación entre la ocasión dos y tres era de 0,978, en el modelo 12.3 es de 0,458, o, por ejemplo, en este mismo modelo la correlación entre la ocasión una y dos es de 0,162 y en el modelo 12.1 era de 0,914.

La principal aportación de los resultados de los modelos de la subsituación 12, es que no siempre que exista una fuerza de relación muy intensa entre el intercepto y la pendiente, de direccionalidad positiva, es esperable un patrón creciente entre la primera ocasión de medida y el resto de mediciones, ya que, ese patrón también depende de los diferentes porcentajes de varianza fijados en los diversos coeficientes aleatorios. Si el porcentaje fijado en la varianza del error intrasujeto, es muy pequeño en comparación con el porcentaje fijado en el estatus, aunque la correlación entre el intercepto y la pendiente sea de 0,9, el patrón que se espera dentro de la primera ocasión de medida es un patrón decreciente.

A continuación, se muestran los resultados de la **subsituación 13** (Tabla VI.34 y Tabla VI.35) en la cual se fijó los mismos porcentajes de varianza que en el modelo 8.1, es decir, un 33,3% en la varianza del error intrasujeto, y lo mismo para la varianza de la pendiente y del intercepto del nivel de estudiantes.

En los modelos de la tabla VI.34, se muestra los resultados de los modelos 13 y 13.1, donde la correlación entre el intercepto y la pendiente se fijó a 0,2 y 0,9, respectivamente.

Tabla VI.34. *Matriz de correlaciones del modelo 13 y del modelo 13.1*

MODELO 13 y 13.1		
EFECTOS FIJOS		MATRIZ DE CORRELACIONES
Coef.		
$\boldsymbol{\beta_{000}}$	**4,6**	
$\boldsymbol{\beta_{100}}$	**0,4**	
EFECTOS ALEATORIOS		
NIVEL 3	$\sigma(u_{00j})$ $\sigma(u_{10j})$	
$\boldsymbol{\sigma(u_{00j})}$	0	
$\boldsymbol{\sigma(u_{10j})}$	0	
NIVEL 2	$\sigma(r_{0ij})$ $\sigma(r_{1ij})$	
$\boldsymbol{\sigma(r_{0ij})}$	**33,3%**	
$\boldsymbol{\sigma(r_{1ij})}$	**33,3%**	
NIVEL 1	$\sigma(e_{tij})$	
$\boldsymbol{\sigma(e_{tij})}$	**33,3%**	
$\boldsymbol{R_{aulas}}$	0	
$\boldsymbol{R_{estudiantes}}$	**0,2/0,9**	

MODELO 13

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,4659**	**0,3932**	**0,3342**	**0,2954**	**0,2736**	**0,2546**	**0,2383**	**0,2286**
Y_{1ij}	0,4659	1	**0,7554**	**0,7519**	**0,7408**	**0,7321**	**0,724**	**0,7142**	**0,711**
Y_{2ij}	0,3932	0,7554	1	0,8759	0,8813	0,8794	0,8783	0,874	0,8725
Y_{3ij}	0,3342	0,7519	0,8759	1	0,932	0,9362	0,9374	0,9371	0,9368
Y_{4ij}	0,2954	0,7408	0,8813	0,932	1	0,9569	0,9591	0,9604	0,9609
Y_{5ij}	0,2736	0,7321	0,8794	0,9362	0,9569	1	0,9708	0,9717	0,9727
Y_{6ij}	0,2546	0,724	0,8783	0,9374	0,9591	0,9708	1	0,9778	0,9795
Y_{7ij}	0,2383	0,7142	0,874	0,9371	0,9604	0,9717	0,9778	1	**0,9832**
Y_{8ij}	0,2286	0,711	0,8725	0,9368	0,9609	0,9727	0,9795	0,9832	1

MODELO 13.1

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,6215**	**0,649**	**0,6522**	**0,6559**	**0,6556**	**0,6558**	**0,6549**	**0,6537**
Y_{1ij}	0,6215	1	0,8454	0,8606	0,869	0,871	0,8746	0,8752	0,8749
Y_{2ij}	0,649	0,8454	1	0,9179	0,9277	0,9327	0,9353	0,9367	0,9368
Y_{3ij}	0,6522	0,8606	0,9179	1	0,95	0,9556	0,9589	0,9603	0,9616
Y_{4ij}	0,6559	0,869	0,9277	0,95	1	0,9665	0,97	0,9722	0,9732
Y_{5ij}	0,6556	0,871	0,9327	0,9556	0,9665	1	0,9758	0,9785	0,9797
Y_{6ij}	0,6558	0,8746	0,9353	0,9589	0,97	0,9758	1	0,9824	0,9837
Y_{7ij}	0,6549	0,8752	0,9367	0,9603	0,9722	0,9785	0,9824	1	**0,9858**
Y_{8ij}	0,6537	0,8749	0,9368	0,9616	0,9732	0,9797	0,9837	0,9858	1

En los resultados del **modelo 13,** tal y como se ha resaltado, las correlaciones entre la primera ocasión de medida y el resto de mediciones, siguen mostrando un patrón decreciente como en el modelo 8.1, donde la correlación entre el intercepto y la pendiente se fijó a 0. Ambos modelos (modelo 8.1 y modelo 13), muestran valores de correlación muy similares, así, por ejemplo, mientras que en el modelo 8.2, la correlación entre la ocasión dos y tres era de 0,702, en el modelo 13 es de 0,755.

En cuanto al **modelo 13.1**, como el porcentaje fijado en el estatus es el mismo que el fijado en la varianza del error intrasujeto, ya no se observa dentro de la primera

ocasión de medida un patrón decreciente, es más constante o incluso con cierta tendencia creciente.

Si se compara este modelo 13.1 con el modelo 8.1, las diferencias entre estos modelos son más marcadas que entre los modelos 13 y 8.1, ya que, en estos últimos modelos las correlaciones entre el intercepto y la pendiente son parecidas. Así, mientras que, por ejemplo, en el modelo 13.1 la correlación entre la ocasión una y nueve es de 0,654, en el modelo 8.1, era de 0,088.

Finalmente, dentro de esta subsituación se muestran los resultados obtenidos cuando se fija una correlación negativa de -0,2 (**modelo 13.2**) y -0,9 (**modelo 13.3**).

Tabla VI.35. *Matriz de correlaciones del modelo 13.2 y del modelo 13.3*

MODELO 13.2 y 13.3

EFECTOS FIJOS			**MATRIZ DE CORRELACIONES**
Coef.			
β_{000}	**4,6**		
β_{100}	**0,4**		
EFECTOS ALEATORIOS			
NIVEL 3	$\sigma(u_{00j})$	$\sigma(u_{10j})$	
$\sigma(u_{00j})$	0		
$\sigma(u_{10j})$		0	
NIVEL 2	$\sigma(r_{0ij})$	$\sigma(r_{1ij})$	
$\sigma(r_{0ij})$	**33,3%**		
$\sigma(r_{1ij})$		**33,3%**	
NIVEL 1	$\sigma(e_{tij})$		
$\sigma(e_{tij})$	**33,3%**		
R_{aulas}	0		
$R_{estudiantes}$	**-0,2/-0,9**		

MODELO 13.2

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,3539**	**0,1896**	**0,1041**	**0,0404**	**0,0093**	**-0,0166**	**-0,0326**	**-0,0458**
Y_{1ij}	0,3539	1	**0,647**	**0,6328**	**0,6074**	**0,5891**	**0,5753**	**0,5677**	**0,5602**
Y_{2ij}	0,1896	0,647	1	0,8389	0,8444	0,8407	0,8379	0,8343	0,8303
Y_{3ij}	0,1041	0,6328	0,8389	1	0,9133	0,9173	0,918	0,9176	0,9159
Y_{4ij}	0,0404	0,6074	0,8444	0,9133	1	0,9472	0,9509	0,953	0,9527
Y_{5ij}	0,0093	0,5891	0,8407	0,9173	0,9472	1	0,9643	0,9677	0,9688
Y_{6ij}	-0,0166	0,5753	0,8379	0,918	0,9509	0,9643	1	0,9751	0,9766
Y_{7ij}	-0,0326	0,5677	0,8343	0,9176	0,953	0,9677	0,9751	1	0,9813
Y_{8ij}	-0,0458	0,5602	0,8303	0,9159	0,9527	0,9688	0,9766	0,9813	1

MODELO 13.3

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,0528**	**-0,3587**	**-0,5029**	**-0,5533**	**-0,5798**	**-0,5931**	**-0,6037**	**-0,6086**
Y_{1ij}	0,0528	**1**	**0,1942**	**0,1634**	**0,1571**	**0,1379**	**0,1345**	**0,1284**	**0,1261**
Y_{2ij}	-0,3587	0,1942	1	0,6809	0,7057	0,7125	0,7161	0,7162	0,7181
Y_{3ij}	-0,5029	0,1634	0,6809	1	0,8616	0,8746	0,8832	0,8855	0,8885
Y_{4ij}	-0,5533	0,1571	0,7057	0,8616	1	0,9251	0,9342	0,9379	0,9403
Y_{5ij}	-0,5798	0,1379	0,7125	0,8746	0,9251	1	0,953	0,9573	0,9608
Y_{6ij}	-0,5931	0,1345	0,7161	0,8832	0,9342	0,953	1	0,9688	0,9718
Y_{7ij}	-0,6037	0,1284	0,7162	0,8855	0,9379	0,9573	0,9688	1	0,9768
Y_{8ij}	-0,6086	0,1261	0,7181	0,8885	0,9403	0,9608	0,9718	0,9768	1

Los resultados de la tabla anterior VI.35 muestran, nuevamente, que el patrón decreciente entre la primera medición y el resto de aplicaciones es más marcado cuando la correlación entre el intercepto y la pendiente es negativa. Al igual que en el modelo 8.1, esa tendencia decreciente se observa hasta la segunda ocasión de medida.

Existe una diferencia muy marcada entre los valores de correlación del **modelo 13.2** y del **modelo 13.3**, ya que, mientras en este último modelo la correlación entre la ocasión dos y tres es de 0,194, en el modelo 13.3 es de 0,647. También existe una discrepancia importante entre fijar una correlación positiva alta de 0,9 (modelo 13.1) y una correlación negativa alta de -0,9 (modelo 13.3), ya que, mientras en este último modelo se necesitan entre cinco o seis medidas para obtener correlaciones superiores a 0,90, en el modelo 13.1, con tres o cuatro ocasiones de medida, se obtienen los mismos valores de correlación.

A la vista de los resultados de los modelos de la subsituación 13 se observa, que cuando el porcentaje de varianza fijado en el error intrasujeto es menor a la suma de los valores de varianza del intercepto y la pendiente de un mismo nivel, el efecto de los valores de correlación fijado entre el intercepto y la pendiente, es más marcado que cuando el porcentaje de varianza del error intrasujeto era superior a la suma de los coeficientes aleatorios del nivel de estudiante. Esta diferencia se evidencia aun más cuando se comparan modelos donde se fija una correlación negativa, con modelos donde se fija una correlación positiva. Además, nuevamente se demuestra que no siempre que se fija una correlación alta y positiva (0,9) entre el intercepto y la pendiente, es esperable que el ERM se mitigue a lo largo de las ocasiones de medida (patrón creciente), sino, que, es posible que se mantenga constante o bien que decrezca, todo ello depende de los porcentajes de varianza fijados en los diferentes coeficientes aleatorios.

Finalmente, dentro de la *situación b*, se muestran los resultados de los modelos de la **subsituación 14** (Tabla VI.36 y VI.37), en los cuales, no se fijó variabilidad en la pendiente del nivel de estudiante. Toda la varianza se encuentra repartida entre el intercepto del nivel de aula (55%) y del nivel de estudiante (15%), la varianza de la pendiente del nivel tres (25%) y la varianza del error intrasujeto (5%).

Concretamente, en la tabla siguiente VI.36 se presentan los resultados de los **modelos 14** y **14.1**, en los cuales la correlación entre el intercepto y la pendiente se fijó a 0,2 y 0,9, respectivamente.

Tabla VI.36. *Matriz de correlaciones del modelo 14 y del modelo 14.1*

MODELO 14 y 14.1		
EFECTOS FIJOS		**MATRIZ DE CORRELACIONES**
Coef.		**MODELO 14**
$\boldsymbol{\beta_{000}}$	**4,6**	
$\boldsymbol{\beta_{100}}$	**0,4**	
EFECTOS ALEATORIOS		
NIVEL 3	$\sigma(u_{00j})$ $\sigma(u_{10j})$	
$\boldsymbol{\sigma(u_{00j})}$	**55%**	
$\boldsymbol{\sigma(u_{10j})}$	**25%**	
NIVEL 2	$\sigma(r_{0ij})$ $\sigma(r_{1ij})$	
$\boldsymbol{\sigma(r_{0ij})}$	**15%**	
$\boldsymbol{\sigma(r_{1ij})}$	**0**	
NIVEL 1	$\sigma(e_{tij})$	
$\boldsymbol{\sigma(e_{tij})}$	**5%**	
$\boldsymbol{R_{aulas}}$	**0,2/0,9**	
$\boldsymbol{R_{estudiantes}}$	0	

MODELO 14

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,8468**	**0,71**	**0,6035**	**0,5322**	**0,482**	**0,4425**	**0,4137**	**0,3913**
Y_{1ij}	0,8468	1	**0,9353**	0,888	0,849	0,8185	0,7933	0,774	0,759
Y_{2ij}	0,71	0,9353	1	**0,9701**	0,9541	0,9381	0,9239	0,9122	0,9026
Y_{3ij}	0,6035	0,888	0,9701	1	0,985	0,9786	0,9711	0,9645	0,9586
Y_{4ij}	0,5322	0,849	0,9541	0,985	1	0,991	0,9876	0,984	0,9804
Y_{5ij}	0,482	0,8185	0,9381	0,9786	0,991	1	0,9939	0,9922	0,9901
Y_{6ij}	0,4425	0,7933	0,9239	0,9711	0,9876	0,9939	1	0,9956	0,9948
Y_{7ij}	0,4137	0,774	0,9122	0,9645	0,984	0,9922	0,9956	1	0,9967
Y_{8ij}	0,3913	0,759	0,9026	0,9586	0,9804	0,9901	0,9948	0,9967	1

MODELO 14.1

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	0,921	0,8934	0,8692	0,8523	0,8401	0,8293	0,8216	0,8152
Y_{1ij}	0,921	1	**0,9702**	0,9625	0,9555	0,9494	0,9441	0,9399	0,9362
Y_{2ij}	0,8934	0,9702	1	0,9843	0,9821	0,9796	0,9766	0,9742	0,9719
Y_{3ij}	0,8692	0,9625	0,9843	1	0,9905	0,9899	0,9886	0,9873	0,9861
Y_{4ij}	0,8523	0,9555	0,9821	0,9905	1	0,9937	0,9934	0,9929	0,9921
Y_{5ij}	0,8401	0,9494	0,9796	0,9899	0,9937	1	0,9954	0,9954	0,995
Y_{6ij}	0,8293	0,9441	0,9766	0,9886	0,9934	0,9954	1	0,9967	0,9966
Y_{7ij}	0,8216	0,9399	0,9742	0,9873	0,9929	0,9954	0,9967	1	0,9973
Y_{8ij}	0,8152	0,9362	0,9719	0,9861	0,9921	0,995	0,9966	0,9973	1

Lo primero que se observa en la tabla anterior VI.36, donde el rendimiento de los estudiantes a lo largo del tiempo se debe al efecto de las aulas y no a sus propias características, es que, ambos modelos muestran un patrón decreciente dentro de la primera ocasión de medida (más pronunciado en el modelo 14). En el modelo 14 este patrón se vuelve más constante a partir de la segunda medición, y a partir de la primera medición, en el modelo 14.1.

En ambos modelos (modelo 14 y modelo 14.1), con un bajo número de mediciones es posible obtener estimaciones fiables del rendimiento de los estudiantes. Concretamente en el modelo 14, con tres, cuatro medidas que es el número de mediciones mínimas requeridas en los modelos de crecimiento, se obtiene una correlación de 0,97. Dentro de estas ocasiones de medida, y también en el resto, se mantiene una ligera tendencia hacia un patrón decreciente.

En el **modelo 14.1**, también se obtiene una correlación de 0,97, pero con menos mediciones que en el modelo 14, ya que la correlación entre la segunda y tercera medición es de 0,972.

En estos dos modelos de la subsituación 14, no merece la pena realizar más de tres mediciones, ya que, la información que aportan el resto de ocasiones de medida no es muy significativa, y, por tanto, no compensa al coste que supone la realización de las mismas.

En la tabla siguiente VI.37, se muestran los resultados obtenidos cuando se fija una correlación negativa de -0,2 (**modelo 14.2**) y de -0,9 (**modelo 14.3**).

Tabla VI.37. *Matriz de correlaciones del modelo 14.2 y del modelo 14.3*

MODELO 14.2 y 14..3		
EFECTOS FIJOS		MATRIZ DE CORRELACIONES
Coef.		
$\boldsymbol{\beta_{000}}$	**4,6**	
$\boldsymbol{\beta_{100}}$	**0,4**	
EFECTOS ALEATORIOS		
NIVEL 3	$\sigma(u_{00j})$ $\sigma(u_{10j})$	
$\boldsymbol{\sigma(u_{00j})}$	**55%**	
$\boldsymbol{\sigma(u_{10j})}$	**25%**	
NIVEL 2	$\sigma(r_{0ij})$ $\sigma(r_{1ij})$	
$\boldsymbol{\sigma(r_{0ij})}$	**15%**	
$\boldsymbol{\sigma(r_{1ij})}$	0	
NIVEL 1	$\sigma(e_{tij})$	
$\boldsymbol{\sigma(e_{tij})}$	**5%**	
$\boldsymbol{R_{aulas}}$	**-0,2/-0,9**	
$\boldsymbol{R_{estudiantes}}$	0	

MODELO 14.2

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,7795**	**0,5234**	**0,3467**	**0,2306**	**0,1567**	**0,1039**	**0,0651**	**0,0364**
Y_{1ij}	0,7795	1	0,8783	0,7859	0,7116	0,6588	0,6195	0,5897	0,5669
Y_{2ij}	0,5234	0,8783	1	**0,9507**	0,9205	0,8931	0,8709	0,8526	0,8381
Y_{3ij}	0,3467	0,7859	0,9507	1	**0,9763**	0,9651	0,9535	0,9431	0,9342
Y_{4ij}	0,2306	0,7116	0,9205	0,9763	1	0,9869	0,9822	0,9768	0,9716
Y_{5ij}	0,1567	0,6588	0,8931	0,9651	0,9869	1	0,9916	0,9894	0,9866
Y_{6ij}	0,1039	0,6195	0,8709	0,9535	0,9822	0,9916	1	0,9943	0,9931
Y_{7ij}	0,0651	0,5897	0,8526	0,9431	0,9768	0,9894	0,9943	1	0,9959
Y_{8ij}	0,0364	0,5669	0,8381	0,9342	0,9716	0,9866	0,9931	0,9959	1

MODELO 14.3

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	**1**	**0,736**	**0,0823**	**-0,3224**	**-0,4857**	**-0,5642**	**-0,6101**	**-0,639**	**-0,6592**
Y_{1ij}	0,736	1	0,5531	0,2211	0,0595	-0,0287	-0,0817	-0,1167	-0,1419
Y_{2ij}	0,0823	0,5531	1	0,8279	0,752	0,7009	0,6658	0,6412	0,623
Y_{3ij}	-0,3224	0,2211	0,8279	1	**0,9465**	0,93	0,9157	0,9039	0,8947
Y_{4ij}	-0,4857	0,0595	0,752	0,9465	1	**0,9766**	0,9722	0,967	0,9623
Y_{5ij}	-0,5642	-0,0287	0,7009	0,93	0,9766	1	0,9869	0,9852	0,9831
Y_{6ij}	-0,6101	-0,0817	0,6658	0,9157	0,9722	0,9869	1	0,9919	0,9913
Y_{7ij}	-0,639	-0,1167	0,6412	0,9039	0,967	0,9852	0,9919	1	0,9942
Y_{8ij}	-0,6592	-0,1419	0,623	0,8947	0,9623	0,9831	0,9913	0,9942	1

En la tabla anterior se muestra que los valores de correlación obtenidos son inferiores a los de los modelos anteriores, y que la tendencia decreciente estudiada con anterioridad es más pronunciada.

En el caso del **modelo 14.2**, hay que destacar el patrón decreciente tan marcado existente dentro de las correlaciones de la primera ocasión de medida y el resto de ocasiones, aunque dicho patrón es todavía mas acentuado en el **modelo 14.3**, ya que, por ejemplo, pasa de un valor de correlación entre la primera ocasión de medida y la segunda de 0,736, a una correlación de 0,082 entre la primera y tercera medición.

En el modelo donde la correlación entre el intercepto y la pendiente se fijó a -0,2 (modelo 14.2), es necesario realizar entre tres y cuatro mediciones para obtener una correlación superior a 0,95 y, cuatro o cinco para conseguir una correlación de 0,97. En cuanto al modelo 14.3, se obtienen valores de correlación cercanos a 0,95 con cuatro o cinco mediciones, y con cinco, seis mediciones, valores de 0,97. Por lo tanto, dicho modelo es el que más mediciones requiere para obtener valores de correlación muy altos.

A la vista de los resultados anteriores, donde no se fija varianza en la pendiente del nivel de estudiante, y la varianza de la pendiente del nivel de aula es de un 25%, los valores de correlación tienen un importante efecto en la fiabilidad de las estimaciones de las puntuaciones de los estudiantes. Cuando la correlación es alta y positiva, es decir, el rendimiento de partida de las aulas, esta muy relacionado con el que alcanzan a lo largo del tiempo, con dos ocasiones de medida las correlaciones ya son superiores a 0,90. Sin embargo, cuando existe una tendencia muy marcada en la cual el rendimiento de partida de las aulas, no esta muy relacionado con el que alcanzan a lo largo del tiempo. (modelo 14.3), es necesario realizar entre cuatro y cinco aplicaciones para obtener correlaciones superiores a 0,90.

VI.1.2.3. Resultados de los modelos de la situación c

Finalmente, dentro de este primer problema de investigación, se muestran los resultados de los modelos desarrollados para dar respuesta al objetivo específico 2 de la ***situación c,*** en la cual se fijó varianza en todos los coeficientes aleatorios y, distintos valores de correlación entre el intercepto y la pendiente, tanto del nivel de aula, como del nivel de estudiante

Los primeros resultados que se presentan son los del **modelo 15**, en el cual, recuérdese que la varianza total del nivel de estudiante es superior a la varianza total del nivel de aula y, dentro de estos niveles se fija un mayor porcentaje de varianza en el intercepto que en la pendiente, tal y como, ocurre en un gran número de investigaciones (Castro et al., 2009; Navarro; 2013; Raudenbush & Bryk, 2002; Zvoch & Stevens, 2003; Zvoch & Stevens, 2008). En cuanto, a la correlación entre el intercepto y la pendiente del nivel de estudiante se fijó a 0,20 y, la del nivel de aula a -0,15. Los resultados de este modelo se presentan en la siguiente Tabla VI.38.

Tabla VI.38. *Matriz de correlaciones del modelo 15*

MODELO 15										
EFECTOS FIJOS			**MATRIZ DE CORRELACIONES**								
Coef.				Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
β_{000}	**4,6**		Y_{0ij}	**1**	**0,621**	**0,4906**	**0,4028**	**0,339**	**0,3001**	**0,2694**	**0,2457**	**0,228**
β_{100}	**0,4**		Y_{1ij}	0,621	1	0,8233	0,7939	0,7662	0,7432	0,726	0,7116	0,7003
EFECTOS ALEATORIOS			Y_{2ij}	0,4906	0,8233	1	**0,915**	0,9084	0,9006	0,8923	0,8848	0,8788
NIVEL 3	$\sigma(u_{00j})$	$\sigma(u_{10j})$	Y_{3ij}	0,4028	0,7939	0,915	**1**	**0,9519**	0,9507	0,9481	0,9442	0,9411
$\sigma(u_{00j})$	**15%**		Y_{4ij}	0,339	0,7662	0,9084	0,9519	**1**	**0,9703**	0,971	0,9698	0,9682
$\sigma(u_{10j})$		**5%**	Y_{5ij}	0,3001	0,7432	0,9006	0,9507	0,9703	1	**0,9792**	**0,9794**	**0,9792**
NIVEL 2	$\sigma(r_{0ij})$	$\sigma(r_{1ij})$	Y_{6ij}	0,2694	0,726	0,8923	0,9481	0,971	0,9792	1	0,9852	0,9856
$\sigma(r_{0ij})$	**35%**		Y_{7ij}	0,2457	0,7116	0,8848	0,9442	0,9698	0,9794	0,9852	1	0,9885
$\sigma(r_{1ij})$		**25%**	Y_{8ij}	0,228	0,7003	0,8788	0,9411	0,9682	0,9792	0,9856	0,9885	1
NIVEL 1	$\sigma(e_{tij})$											
$\sigma(e_{tij})$	**20%**											
R_{aulas}	**-0,15**											
$R_{estudiantes}$	**0,20**											

Lo primero que se aprecia en el modelo 15, en el cual, por primera vez, se considera que en general, los estudiantes parten de distintos niveles de rendimiento académico y crecen a distinto ritmo, por ello, en promedio, dentro de una misma aula los estudiantes son distintos, tanto en el nivel de partida, como en la pendiente de crecimiento y, el rendimiento entre las aulas también es diferente, debido al efecto de las características sistemáticas de los estudiantes y de las propias aulas, es que existe un patrón decreciente entre la primera ocasión de medida (valores resaltados en negrita) y el resto de ocasiones, dentro del resto de mediciones dicha tendencia es mas constante.

Se observa en los valores destacados, que con tres o cuatro ocasiones de medida se obtienen valores de correlación superiores a 0,90, con cuatro o cinco mediciones superiores a 0,95, y, con más de cinco o seis son superiores a 0,97. Por lo tanto, en este modelo, elaborado a partir de los resultados obtenidos en la investigación educativa (Castro et al., 2009; Navarro; 2013; Raudenbush & Bryk, 2002; Zvoch & Stevens, 2003; Zvoch & Stevens, 2008), con tres o cuatro medidas, que son el mínimo de mediciones requeridas en los modelos longitudinales de crecimiento (Raudenbush & Bryk, 2002), se pueden obtener estimaciones fiables del rendimiento de los estudiantes, ya que, se generan correlaciones superiores a 0,90. Pero, si se desea una mayor precisión, superior a 0,95, con el fin de que exista un menor margen de error en la medición de los niveles de logro de los estudiantes, es necesario realizar al menos cuatro o cinco medidas. A partir de cinco o seis

aplicaciones, no se aporta mucha información extra, ya que, las correlaciones no son muy superiores a 0,97.

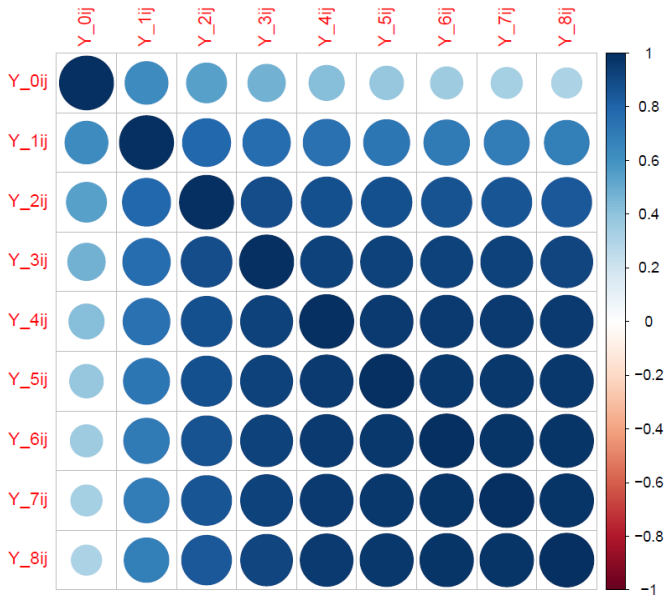
El modelo que se presenta a continuación, es el modelo generado a partir de los resultados obtenidos en el modelo multinivel y longitudinal, de Raudenbush y Bryk (2002), para la quinta ocasión de medida. En este **modelo 16**, tal y como se comentó en el procedimiento de generación de modelos, la suma total de las varianzas originales no son el 100%, sino, un 77%, ya que no se tiene en cuenta el valor de las covarianzas, sino de la correlación. Aun así, en la tabla siguiente (VI.39), entre paréntesis, con el fin de facilitar su interpretación, aparecen los porcentajes correspondientes al considerar que los mismos suman el 100%.

Tabla VI.39. *Matriz de correlaciones del modelo 16*

MODELO 16		
EFECTOS FIJOS		
Coef.		
β_{000}	**4,6**	
β_{100}	**0,4**	
EFECTOS ALEATORIOS		
NIVEL 3	$\sigma(u_{00j})$	$\sigma(u_{10j})$
$\sigma(u_{00j})$	**3,36% (4,35%)**	
$\sigma(u_{10j})$		**9,45% (12,27%)**
NIVEL 2	$\sigma(r_{0ij})$	$\sigma(r_{10j})$
$\sigma(r_{0ij})$	**37,7% (48,96%)**	
$\sigma(r_{10j})$		**6,49% (8,42%)**
NIVEL 1	$\sigma(e_{tij})$	
$\sigma(e_{tij})$	**20% (26%)**	
R_{aulas}	**0,36**	
$R_{estudiantes}$	**0,24**	

MATRIZ DE CORRELACIONES

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,6287**	**0,5468**	**0,4779**	**0,4236**	**0,3857**	**0,3554**	**0,3313**	**0,313**
Y_{1ij}	0,6287	1	0,7898	0,7692	0,7481	0,725	0,7086	0,6938	0,682
Y_{2ij}	0,5468	0,7898	1	0,8803	0,8773	0,8703	0,8624	0,8546	0,8483
Y_{3ij}	0,4779	0,7692	0,8803	1	**0,9263**	0,9264	0,9251	0,9211	0,9181
Y_{4ij}	0,4236	0,7481	0,8773	0,9263	1	0,9514	0,9537	0,9529	0,9513
Y_{5ij}	0,3857	0,725	0,8703	0,9264	0,9514	1	0,965	0,9659	0,9664
Y_{6ij}	0,3554	0,7086	0,8624	0,9251	0,9537	0,965	1	0,9747	0,9757
Y_{7ij}	0,3313	0,6938	0,8546	0,9211	0,9529	0,9659	0,9747	1	0,9799
Y_{8ij}	0,313	0,682	0,8483	0,9181	0,9513	0,9664	0,9757	0,9799	1



Lo primero que se observa en los resultados del modelo 16, es que, el patrón de correlación entre la primera ocasión de medida con el resto de ocasiones es decreciente, ese patrón algo menos pronunciado se extiende a otras ocasiones como por ejemplo la ocasión dos y tres.

En cuanto a la fiabilidad de las estimaciones en las puntuaciones de los estudiantes, en este modelo con cuatro o cinco aplicaciones, ya se puede predecir el rendimiento de los estudiantes con cierta precisión, ya que, las correlaciones entre las diferentes ocasiones de medida son superiores a 0,90.

Como la varianza del intercepto supone un 41% de la varianza total del modelo (sobre el 100%), se parte de una correlación moderada de 0,629, que va decreciendo a medida que se añaden mediciones. Otra razón que hace que en general en la matriz de correlaciones no se obtengan valores entre las puntuaciones de los estudiantes bajos, a excepción del patrón decreciente entre la primera ocasión de medida y el resto de ocasiones, se debe a que el porcentaje de varianza del error intrasujeto es un 26% de la varianza total del modelo (sobre el 100%), el resto de varianza se encuentra repartido entre los coeficientes aleatorios (74%).

Finalmente, en cuanto a los valores de correlación fijados entre el intercepto y la pendiente del nivel de estudiante, y del nivel de aula, son valores con una direccionalidad positiva, es decir, los estudiantes o aulas que parten de niveles de rendimiento alto crecen mucho a lo largo del tiempo y los que parten de rendimiento bajo, el mismo no se incrementa excesivamente a medida que se añaden aplicaciones. En cuanto a la fuerza de relación entre el intercepto y la pendiente, es moderada siendo algo superior en el nivel de aula (0,34), por lo tanto, en este nivel existe más relación entre el rendimiento del que parten las aulas y el que alcanzan a lo largo del tiempo, que, en el nivel de estudiante, donde la correlación se fijó a 0,24.

Como en ambos niveles los valores de correlación no son elevados, se sigue reflejando el mismo patrón decreciente entre la primera ocasión de medida y el resto de ocasiones, que se observaba en los modelos donde la correlación entre el intercepto y la pendiente se fijaba a 0,2 0 o -0,2. Por lo tanto, dicho patrón decreciente siguiendo a Nesselroade et al., (1980), es un indicativo de que a medida que se añaden aplicaciones el efecto de regresión a la media se incrementa.

En los resultados de los siguientes modelos de la *situación c*, al fijar dos valores positivos y dos valores negativos en la correlación entre el intercepto y la pendiente del nivel de aula y del nivel de estudiante, cada modelo cuenta con 16 combinaciones distintas de correlaciones, tal y como, se explicó en la Tabla V.20

Como se comentó en el apartado *procedimiento de generación de modelos (V.3.3.3),* por cuestiones de espacio no en todos los modelos se presentan las 16 combinaciones, solo aquellas que se consideran más relevantes, aunque los resultados del resto de combinaciones pueden ser consultados en el Anexo III.

En los modelos en los cuales se presentan los resultados de las 16 combinaciones, la matriz de datos de los cuatro modelos donde la correlación fijada entre el intercepto y la pendiente en el nivel de aula es siempre igual aparecen en una misma tabla, y lo que cambia, es la correlación entre el estatus y la pendiente del nivel de estudiante.

A continuación, se muestran los resultados de la **subsituación 17** (Tabla VI.40, Tabla VI.41, Tabla VI.42 y Tabla VI.43), donde los porcentajes fijados fueron los del modelo de Raudenbush y Bryk (2002), considerando que la suma de las varianzas es el 100%. Dentro de esta situación, se desarrollarón 16 modelos que hacen referencia cada una de las 16 combinaciones explicadas anteriormente en la Tabla V.20.

En la tabla siguiente (Tabla VI. 40) se muestra los resultados de los modelos de la **combinación 1, 2, 3 y 4,** donde la correlación entre el intercepto y la pendiente del nivel de aula se fijó a 0,2.

Tabla VI.40. *Matriz de correlaciones de las combinaciones 1, 2, 3 y 4 de la subsituación 17*

SUBSITUACIÓN 17

EFECTOS FIJOS	Coef.
β_{000}	4,6
β_{100}	0,4
EFECTOS ALEATORIOS	
NIVEL 3	$\sigma(u_{00j})$ $\sigma(u_{10j})$
$\sigma(u_{00j})$	4,35%
$\sigma(u_{10j})$	12,27%
NIVEL 2	$\sigma(r_{0ij})$ $\sigma(r_{1ij})$
$\sigma(r_{0ij})$	48,96%
$\sigma(r_{1ij})$	8,42%
NIVEL 1	$\sigma(e_{tij})$
$\sigma(e_{tij})$	26%

Raulas y Restudiantes

COMBINACIÓN 1

R_{aulas}: 0,2

$R_{estudiantes}$:0,2

COMBINACIÓN 2

R_{aulas}: 0,2

$R_{estudiantes}$:0,9

COMBINACIÓN 3

R_{aulas}: 0,2

$R_{estudiantes}$: -0,2

COMBINACIÓN 4

R_{aulas}: 0,2

$R_{estudiantes}$: -0,9

MATRIZ DE CORRELACIONES

COMBINACIÓN 1

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,6152**	**0,5134**	**0,4334**	**0,3712**	**0,3284**	**0,2957**	**0,2744**	**0,2553**
Y_{1ij}	0,6152	1	**0,7772**	**0,7485**	**0,7258**	**0,7021**	**0,68**	**0,6661**	**0,6542**
Y_{2ij}	0,5134	0,7772	1	**0,8724**	**0,8699**	**0,8611**	**0,8516**	**0,8443**	**0,8382**
Y_{3ij}	0,4334	0,7485	0,8724	1	0,9229	0,9238	0,9203	0,9184	0,9149
Y_{4ij}	0,3712	0,7258	0,8699	0,9229	1	0,9498	0,9508	0,9508	0,9498
Y_{5ij}	0,3284	0,7021	0,8611	0,9238	0,9498	1	0,9645	0,9669	0,9668
Y_{6ij}	0,2957	0,68	0,8516	0,9203	0,9508	0,9645	1	0,9738	0,9749
Y_{7ij}	0,2744	0,6661	0,8443	0,9184	0,9508	0,9669	0,9738	1	0,9801
Y_{8ij}	0,2553	0,6542	0,8382	0,9149	0,9498	0,9668	0,9749	0,9801	1

COMBINACIÓN 2

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,6887**	**0,6666**	**0,6342**	**0,617**	**0,5953**	**0,5787**	**0,5679**	**0,5584**
Y_{1ij}	0,6887	1	**0,8409**	**0,84**	**0,8373**	**0,8271**	**0,82**	**0,8142**	**0,8094**
Y_{2ij}	0,6666	0,8409	1	0,9095	0,9139	0,9134	0,9094	0,9071	0,9049
Y_{3ij}	0,6342	0,84	0,9095	1	0,943	0,9465	0,946	0,9451	0,9443
Y_{4ij}	0,617	0,8373	0,9139	0,943	1	0,9608	0,9629	0,9636	0,9638
Y_{5ij}	0,5953	0,8271	0,9134	0,9465	0,9608	1	0,9714	0,9735	0,974
Y_{6ij}	0,5787	0,82	0,9094	0,946	0,9629	0,9714	1	0,9781	0,9797
Y_{7ij}	0,5679	0,8142	0,9071	0,9451	0,9636	0,9735	0,9781	1	0,9832
Y_{8ij}	0,5584	0,8094	0,9049	0,9443	0,9638	0,974	0,9797	0,9832	1

COMBINACIÓN 3

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,5966**	**0,4533**	**0,3306**	**0,2496**	**0,1927**	**0,1509**	**0,1204**	**0,0992**
Y_{1ij}	0,5966	1	**0,7171**	**0,6605**	**0,6196**	**0,5836**	**0,5545**	**0,5332**	**0,5178**
Y_{2ij}	0,4533	0,7171	1	**0,8352**	**0,8259**	**0,811**	**0,7945**	**0,7818**	**0,7723**
Y_{3ij}	0,3306	0,6605	0,8352	1	**0,9037**	**0,9033**	**0,8991**	**0,8928**	**0,8877**
Y_{4ij}	0,2496	0,6196	0,8259	0,9037	1	0,9355	0,9378	0,9362	0,9337
Y_{5ij}	0,1927	0,5836	0,811	0,9033	0,9355	1	0,9555	0,958	0,957
Y_{6ij}	0,1509	0,5545	0,7945	0,8991	0,9378	0,9555	1	0,9687	0,9701
Y_{7ij}	0,1204	0,5332	0,7818	0,8928	0,9362	0,958	0,9687	1	0,976
Y_{8ij}	0,0992	0,5178	0,7723	0,8877	0,9337	0,957	0,9701	0,976	1

COMBINACIÓN 4

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,5046**	**0,2302**	**0,0429**	**-0,0749**	**-0,1524**	**-0,1922**	**-0,2254**	**-0,2478**
Y_{1ij}	0,5046	1	**0,5827**	**0,4836**	**0,4125**	**0,3555**	**0,3239**	**0,297**	**0,2754**
Y_{2ij}	0,2302	0,5827	1	**0,7705**	**0,7462**	**0,7221**	**0,7094**	**0,691**	**0,6783**
Y_{3ij}	0,0429	0,4836	0,7705	1	**0,8714**	**0,8706**	**0,8658**	**0,8589**	**0,8533**
Y_{4ij}	-0,0749	0,4125	0,7462	0,8714	1	0,9222	0,9251	0,9237	0,9216
Y_{5ij}	-0,1524	0,3555	0,7221	0,8706	0,9222	1	0,9504	0,9525	0,9532
Y_{6ij}	-0,1922	0,3239	0,7094	0,8658	0,9251	0,9504	1	0,965	0,9672
Y_{7ij}	-0,2254	0,297	0,691	0,8589	0,9237	0,9525	0,965	1	0,9742
Y_{8ij}	-0,2478	0,2754	0,6783	0,8533	0,9216	0,9532	0,9672	0,9742	1

Lo primero que se observa en los resultados de los modelos anteriores, es que, la correlación entre la primera ocasión de medida y el resto de ocasiones tiende a un patrón decreciente; este patrón se extiende dentro de otras ocasiones de medida, dependiendo de la correlación fijada entre el intercepto y la pendiente del nivel de estudiante. Así, por ejemplo, el patrón autorregresivo es más pronunciado y se extiende dentro de un mayor número de ocasiones, en los modelos donde se fijan valores de correlación negativos entre el intercepto y la pendiente del nivel de estudiante (combinación 3 y 4). Cuando existe una relación positiva, pero no muy intensa, entre el rendimiento del que parten las aulas y los estudiantes, y el que adquieren a lo largo del tiempo (combinación 1), también ese patrón decreciente se extiende más allá de la segunda medición. Sin embargo, cuando a nivel de aula no hay mucha relación entre el estatus y la pendiente, pero si entre el intercepto y la pendiente de los estudiantes, siendo dicha relación positiva (combinación 2), el patrón decreciente no es tan pronunciado como en el resto de modelos.

Como es esperable, los valores de correlación más bajos de las cuatro combinaciones presentadas en la tabla anterior, se obtienen en la combinación 4, ya que, es la combinación donde se fijó, a nivel de estudiante un valor de correlación entre el intercepto y la pendiente alto y negativo (-0,9), en este modelo se requieren entre cinco y seis medidas para obtener correlaciones superiores a 0,90. Por el contrario, es en la combinación 2, donde se fijó ese mismo valor en la correlación entre el intercepto y la pendiente a nivel de estudiante, pero con direccionalidad positiva (0,9), cuando se obtienen las correlaciones más altas con menos aplicaciones. En este modelo con tres o cuatro mediciones se obtienen estimaciones bastante fiables sobre las puntuaciones de los estudiantes, ya que, las correlaciones son superiores a 0,90. En cuanto al resto de combinaciones con cuatro – cinco mediciones la fiabilidad con la que se estima las puntuaciones de los estudiantes es alta (correlaciones en torno a 0,90).

A continuación, se muestran los resultados de los modelos de la **subsituación 17**, correspondientes a las **combinaciones 5, 6, 7 y 8.**

Tabla VI.41. *Matriz de correlaciones de las combinaciones 5, 6, 7 y 8 de la subsituación 17*

SUBSITUACIÓN 17

EFECTOS FIJOS	Coef.
β_{000}	4,6
β_{100}	0,4
EFECTOS ALEATORIOS	
NIVEL 3	$\sigma(u_{00j})$ $\sigma(u_{10j})$
$\sigma(u_{00j})$	4,35%
$\sigma(u_{10j})$	12,27%
NIVEL 2	$\sigma(r_{0ij})$ $\sigma(r_{1ij})$
$\sigma(r_{0ij})$	48,96%
$\sigma(r_{1ij})$	8,42%
NIVEL 1	$\sigma(e_{tij})$
$\sigma(e_{tij})$	26%

Raulas y Restudiantes

COMBINACIÓN 5

R_{aulas}: 0,9

$R_{estudiantes}$:0,2

COMBINACIÓN 6

R_{aulas}: 0,9

$R_{estudiantes}$:0,9

COMBINACIÓN 7

R_{aulas}: 0,9

$R_{estudiantes}$: -0,2

COMBINACIÓN 8

R_{aulas}: 0,9

$R_{estudiantes}$: -0,9

MATRIZ DE CORRELACIONES

COMBINACIÓN 5

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,6431**	**0,572**	**0,5181**	**0,4725**	**0,4418**	**0,4154**	**0,3961**	**0,3809**
Y_{1ij}	0,6431	1	**0,7986**	**0,7906**	**0,7702**	**0,7553**	**0,7396**	**0,7304**	**0,7201**
Y_{2ij}	0,572	0,7986	1	0,8856	0,8836	0,8797	0,8733	0,8686	0,862
Y_{3ij}	0,5181	0,7906	0,8856	1	0,9287	0,93	0,9293	0,9281	0,9241
Y_{4ij}	0,4725	0,7702	0,8836	0,9287	1	0,9522	0,9544	0,9551	0,954
Y_{5ij}	0,4418	0,7553	0,8797	0,93	0,9522	1	0,966	0,9676	0,9674
Y_{6ij}	0,4154	0,7396	0,8733	0,9293	0,9544	0,966	1	0,9754	0,9765
Y_{7ij}	0,3961	0,7304	0,8686	0,9281	0,9551	0,9676	0,9754	1	0,9806
Y_{8ij}	0,3809	0,7201	0,862	0,9241	0,954	0,9674	0,9765	0,9806	1

COMBINACIÓN 6

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,7213**	**0,7203**	**0,7152**	**0,6985**	**0,6943**	**0,6852**	**0,6792**	**0,6745**
Y_{1ij}	0,7213	1	0,8539	0,8631	0,8617	0,8608	0,8582	0,8559	0,8534
Y_{2ij}	0,7203	0,8539	1	0,9149	0,9189	0,922	0,9201	0,92	0,9193
Y_{3ij}	0,7152	0,8631	0,9149	1	0,9437	0,9478	0,9491	0,9497	0,9496
Y_{4ij}	0,6985	0,8617	0,9189	0,9437	1	0,9613	0,9636	0,9656	0,9669
Y_{5ij}	0,6943	0,8608	0,922	0,9478	0,9613	1	0,9714	0,9738	0,9751
Y_{6ij}	0,6852	0,8582	0,9201	0,9491	0,9636	0,9714	1	0,9781	0,9801
Y_{7ij}	0,6792	0,8559	0,92	0,9497	0,9656	0,9738	0,9781	1	0,9832
Y_{8ij}	0,6745	0,8534	0,9193	0,9496	0,9669	0,9751	0,9801	0,9832	1

COMBINACIÓN 7

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,6178**	**0,5061**	**0,4146**	**0,352**	**0,3098**	**0,2771**	**0,2499**	**0,2322**
Y_{1ij}	0,6178	1	**0,7592**	**0,7308**	**0,7002**	**0,6737**	**0,6547**	**0,6377**	**0,6248**
Y_{2ij}	0,5061	0,7592	1	**0,8663**	**0,8626**	**0,853**	**0,8424**	**0,833**	**0,8271**
Y_{3ij}	0,4146	0,7308	0,8663	1	0,9218	0,921	0,9205	0,9144	0,9113
Y_{4ij}	0,352	0,7002	0,8626	0,9218	1	0,9489	0,9503	0,949	0,9477
Y_{5ij}	0,3098	0,6737	0,853	0,921	0,9489	1	0,9643	0,965	0,9655
Y_{6ij}	0,2771	0,6547	0,8424	0,9205	0,9503	0,9643	1	0,9738	0,9753
Y_{7ij}	0,2499	0,6377	0,833	0,9144	0,949	0,965	0,9738	1	0,9798
Y_{8ij}	0,2322	0,6248	0,8271	0,9113	0,9477	0,9655	0,9753	0,9798	1

COMBINACIÓN 8

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,5341**	**0,3182**	**0,1599**	**0,0666**	**-0,0022**	**-0,0478**	**-0,0729**	**-0,1004**
Y_{1ij}	0,5341	1	**0,6488**	**0,5755**	**0,5148**	**0,4722**	**0,4409**	**0,4226**	**0,3977**
Y_{2ij}	0,3182	0,6488	1	**0,8073**	**0,7969**	**0,7778**	**0,7628**	**0,7513**	**0,7375**
Y_{3ij}	0,1599	0,5755	0,8073	1	**0,8937**	**0,8934**	**0,8889**	**0,8857**	**0,8781**
Y_{4ij}	0,0666	0,5148	0,7969	0,8937	1	0,9329	0,9351	0,9357	0,9326
Y_{5ij}	-0,0022	0,4722	0,7778	0,8934	0,9329	1	0,9555	0,9585	0,9573
Y_{6ij}	-0,0478	0,4409	0,7628	0,8889	0,9351	0,9555	1	0,9694	0,9708
Y_{7ij}	-0,0729	0,4226	0,7513	0,8857	0,9357	0,9585	0,9694	1	0,9767
Y_{8ij}	-0,1004	0,3977	0,7375	0,8781	0,9326	0,9573	0,9708	0,9767	1

Nuevamente, en los resultados de dichos modelos, se vuelve a observar un patrón decreciente entre la ocasión una y el resto de ocasiones, que en determinadas combinaciones va más allá de la primera medición (sobre todo, en la combinación 7 y 8). En la tabla anterior VI.41 hay que destacar que cuando, tanto a nivel de estudiante como a nivel de aula, el rendimiento de partida, está muy relacionado con el que adquieren a lo largo del tiempo, existiendo una relación positiva entre ambos coeficientes (combinación 2), el patrón tiende a ser decreciente, al igual que ocurria en el modelo 12.1, en el cual se fijó una correlación entre el intercepto y la pendiente de 0,9, y el porcentaje de varianza fijado en el error intrasujeto (5%) era muy inferior al fijado en el estatus (47,5%).

En la combinación 5, se han fijado los mismos valores de correlación que en la combinación 2, pero con el orden de correlación invertido entre el nivel dos y tres, pues, mientras que en la combinación 2 se fijó a nivel de aula una correlación entre el intercepto y la pendiente de 0,2 y a nivel de estudiante de 0,9, en la combinación 5 ocurre lo contrario, se fijó a nivel de aula una correlación de 0,9 y a nivel de estudiante de 0,2. Si se comparan los resultados de ambos modelos, las correlaciones son muy similares, por ejemplo, en dicha combinación 2, la correlación entre la primera ocasión y la segunda medición era de 0,689 y en la combinación 5 es de 0,643.

El modelo de la combinación 6, muestra los valores de correlación más altos estudiados hasta el momento, ya que, la fuerza de asociación entre el intercepto y la pendiente en ambos niveles (aula y estudiante), es de una intensidad alta y de direccionalidad positiva (0,9). En este modelo con tres ocasiones de medida es posible obtener estimaciones fiables del rendimiento de los estudiantes, ya que las correlaciones son superiores a 0,90.

Los siguientes resultados aluden a los modelos de las **combinaciones 9, 10, 11 y 12,** en las cuales la correlación entre el intercepto y la pendiente del nivel tres (aula) se fijó siempre a -0,2.

Tabla VI.42. *Matriz de correlaciones de las combinaciones 9, 10, 11 y 12 de la subsituación 17*

SUBSITUACIÓN 17

EFECTOS FIJOS	Coef.
β_{000}	4,6
β_{100}	0,4
EFECTOS ALEATORIOS	
NIVEL 3	$\sigma(u_{00j})$ $\sigma(u_{10j})$
$\sigma(u_{00j})$	4,35%
$\sigma(u_{10j})$	12,27%
NIVEL 2	$\sigma(r_{0ij})$ $\sigma(r_{1ij})$
$\sigma(r_{0ij})$	48,96%
$\sigma(r_{1ij})$	8,42%
NIVEL 1	$\sigma(e_{tij})$
$\sigma(e_{tij})$	26%

Raulas y Restudiantes

COMBINACIÓN 9

R_{aulas}: -0,2

$R_{estudiantes}$:0,2

COMBINACIÓN 10

R_{aulas}: -0,2

$R_{estudiantes}$:0,9

COMBINACIÓN 11

R_{aulas}: -0,2

$R_{estudiantes}$: -0,2

COMBINACIÓN 12

R_{aulas}: -0,2

$R_{estudiantes}$: -0,9

MATRIZ DE CORRELACIONES

COMBINACIÓN 9

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,6018**	**0,4871**	**0,3952**	**0,333**	**0,2836**	**0,2477**	**0,225**	**0,203**
Y_{1ij}	0,6018	1	**0,7521**	**0,722**	**0,6927**	**0,6643**	**0,6413**	**0,626**	**0,612**
Y_{2ij}	0,4871	0,7521	1	**0,8626**	**0,858**	**0,8458**	**0,8355**	**0,8267**	**0,8193**
Y_{3ij}	0,3952	0,722	0,8626	1	0,9183	0,918	0,9148	0,9119	0,9074
Y_{4ij}	0,333	0,6927	0,858	0,9183	1	0,9458	0,947	0,9474	0,9458
Y_{5ij}	0,2836	0,6643	0,8458	0,918	0,9458	1	0,9624	0,9632	0,9638
Y_{6ij}	0,2477	0,6413	0,8355	0,9148	0,947	0,9624	1	0,9722	0,974
Y_{7ij}	0,225	0,626	0,8267	0,9119	0,9474	0,9632	0,9722	1	0,9785
Y_{8ij}	0,203	0,612	0,8193	0,9074	0,9458	0,9638	0,974	0,9785	1

COMBINACIÓN 10

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,6895**	**0,6486**	**0,6092**	**0,5864**	**0,5646**	**0,5484**	**0,5346**	**0,5254**
Y_{1ij}	0,6895	1	**0,8346**	**0,8307**	**0,8232**	**0,8132**	**0,8077**	**0,7998**	**0,794**
Y_{2ij}	0,6486	0,8346	1	0,9066	0,909	0,908	0,905	0,9027	0,8994
Y_{3ij}	0,6092	0,8307	0,9066	1	0,9421	0,9455	0,946	0,9454	0,9436
Y_{4ij}	0,5864	0,8232	0,909	0,9421	1	0,9599	0,9618	0,9632	0,9623
Y_{5ij}	0,5646	0,8132	0,908	0,9455	0,9599	1	0,9713	0,9735	0,9741
Y_{6ij}	0,5484	0,8077	0,905	0,946	0,9618	0,9713	1	0,9784	0,9797
Y_{7ij}	0,5346	0,7998	0,9027	0,9454	0,9632	0,9735	0,9784	1	0,9834
Y_{8ij}	0,5254	0,794	0,8994	0,9436	0,9623	0,9741	0,9797	0,9834	1

COMBINACIÓN 11

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,5802**	**0,4058**	**0,2813**	**0,1894**	**0,1282**	**0,0879**	**0,0599**	**0,0358**
Y_{1ij}	0,5802	1	**0,6956**	**0,6494**	**0,6031**	**0,5641**	**0,5359**	**0,5137**	**0,4971**
Y_{2ij}	0,4058	0,6956	1	**0,8293**	**0,8153**	**0,801**	**0,7853**	**0,7735**	**0,7631**
Y_{3ij}	0,2813	0,6494	0,8293	1	0,8998	0,8977	0,8932	0,8872	0,882
Y_{4ij}	0,1894	0,6031	0,8153	0,8998	1	0,9357	0,9388	0,9369	0,9353
Y_{5ij}	0,1282	0,5641	0,801	0,8977	0,9357	1	0,9565	0,9591	0,9587
Y_{6ij}	0,0879	0,5359	0,7853	0,8932	0,9388	0,9565	1	0,9682	0,9699
Y_{7ij}	0,0599	0,5137	0,7735	0,8872	0,9369	0,9591	0,9682	1	0,9767
Y_{8ij}	0,0358	0,4971	0,7631	0,882	0,9353	0,9587	0,9699	0,9767	1

COMBINACIÓN 12

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,4769**	**0,1373**	**-0,0744**	**-0,2011**	**-0,2789**	**-0,3254**	**-0,3543**	**-0,378**
Y_{1ij}	0,4769	1	**0,4959**	**0,3887**	**0,3058**	**0,2502**	**0,2082**	**0,183**	**0,1639**
Y_{2ij}	0,1373	0,4959	1	**0,724**	**0,7027**	**0,6816**	**0,6639**	**0,6476**	**0,6356**
Y_{3ij}	-0,0744	0,3887	0,724	1	0,8512	0,8544	0,8493	0,8433	0,837
Y_{4ij}	-0,2011	0,3058	0,7027	0,8512	1	0,9144	0,9182	0,9175	0,9153
Y_{5ij}	-0,2789	0,2502	0,6816	0,8544	0,9144	1	0,9459	0,9493	0,9499
Y_{6ij}	-0,3254	0,2082	0,6639	0,8493	0,9182	0,9459	1	0,9635	0,9649
Y_{7ij}	-0,3543	0,183	0,6476	0,8433	0,9175	0,9493	0,9635	1	0,9733
Y_{8ij}	-0,378	0,1639	0,6356	0,837	0,9153	0,9499	0,9649	0,9733	1

En estos modelos, donde existe una ligera tendencia en la cual, las aulas que parten de niveles de logro bajo, dicho rendimiento se incrementa a lo largo del tiempo y las que parten de niveles de logro alto, se mantienen constante o su rendimiento no crece mucho, se observa una predisposición hacia un patrón decreciente entre la ocasión primera y el resto de ocasiones, que se extiende a otras mediciones.

Entre los resultados de los modelos de estas últimas cuatro combinaciones, las estimaciones son más fiables, con menos ocasiones de medida, en el modelo de la combinación 10. En dicha combinación, el rendimiento del que parten los estudiantes, esta muy relacionado con el que alcanzan a lo largo de las ocasiones de medida; además, los estudiantes que parten de un alto rendimiento su pendiente de crecimiento se incrementa mucho a lo largo del tiempo, mientras que los que parten de bajo rendimiento sus niveles de logro no se incrementa tanto (la correlación entre el intercepto y la pendiente del nivel de estudiante se fijó a 0,9). En este modelo, se fijaron los mismos valores de correlación entre el intercepto y la pendiente que los de la combinación 7, solo que en este último modelo la correlación entre el estatus y la pendiente era menor en el nivel de estudiante (-0,2), que en el nivel de aula (0,9). Si se comparan los resultados de ambos modelos, son muy parecidos, aunque son ligeramente superiores en el modelo de la combinación 10, ya que, por ejemplo, mientras en dicho modelo la correlación entre la ocasión dos y tres es de 0,835, en el modelo de la combinación 7 era de 0,759.

El modelo de la combinación 9, también presenta los mismos valores de correlación entre el intercepto y la pendiente que la combinación 3, pero en este último modelo la correlación entre el intercepto y la pendiente es mayor en el nivel de aula, mientras que en el modelo de la combinación 9, la correlación superior se fijó en el nivel de estudiantes. Ambos modelos muestran correlaciones muy similares, ya que, por ejemplo, la correlación entre la ocasión dos y tres del modelo de la combinación 9 es de 0,752 y en el modelo de la combinación 3, era de 0,717, además, en ambos modelos con cuatro o cinco medidas se obtienen estimaciones bastante fiables.

Finalmente, dentro de la **subsituación 17,** se presentan los resultados de las cuatro últimas combinaciones (de la 13 a la 16).

Tabla VI.43. *Matriz de correlaciones de las combinaciones 13, 14, 15 y 16 de la subsituación 17*

SUBSITUACIÓN 17

EFECTOS FIJOS	Coef.
β_{000}	4,6
β_{100}	0,4
EFECTOS ALEATORIOS	
NIVEL 3	$\sigma(u_{00j})$ $\sigma(u_{10j})$
$\sigma(u_{00j})$	4,35%
$\sigma(u_{10j})$	12,27%
NIVEL 2	$\sigma(r_{0ij})$ $\sigma(r_{1ij})$
$\sigma(r_{0ij})$	48,96%
$\sigma(r_{1ij})$	8,42%
NIVEL 1	$\sigma(e_{tij})$
$\sigma(e_{tij})$	26%

Raulas y Restudiantes

COMBINACIÓN 13

R_{aulas}: -0,9

$R_{estudiantes}$:0,2

COMBINACIÓN 14

R_{aulas}: -0,9

$R_{estudiantes}$:0,9

COMBINACIÓN 15

R_{aulas}: -0,9

$R_{estudiantes}$: -0,2

COMBINACIÓN 16

R_{aulas}: -0,9

$R_{estudiantes}$: -0,9

MATRIZ DE CORRELACIONES

COMBINACIÓN 13

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,5812**	**0,4254**	**0,2968**	**0,2168**	**0,1637**	**0,1181**	**0,0933**	**0,0679**
Y_{1ij}	0,5812	1	**0,7247**	**0,6745**	**0,6327**	**0,5991**	**0,5687**	**0,5503**	**0,5334**
Y_{2ij}	0,4254	0,7247	1	**0,8444**	**0,8376**	**0,8218**	**0,8074**	**0,7974**	**0,7867**
Y_{3ij}	0,2968	0,6745	0,8444	1	0,913	0,912	0,9064	0,9032	0,8974
Y_{4ij}	0,2168	0,6327	0,8376	0,913	1	0,9454	0,9447	0,9446	0,9417
Y_{5ij}	0,1637	0,5991	0,8218	0,912	0,9454	1	0,9626	0,963	0,9621
Y_{6ij}	0,1181	0,5687	0,8074	0,9064	0,9447	0,9626	1	0,9725	0,9737
Y_{7ij}	0,0933	0,5503	0,7974	0,9032	0,9446	0,963	0,9725	1	0,979
Y_{8ij}	0,0679	0,5334	0,7867	0,8974	0,9417	0,9621	0,9737	0,979	1

COMBINACIÓN 14

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,6644**	**0,6087**	**0,5541**	**0,5162**	**0,4899**	**0,464**	**0,4502**	**0,4334**
Y_{1ij}	0,6644	1	**0,8168**	**0,8041**	**0,7898**	**0,7758**	**0,7628**	**0,7547**	**0,7435**
Y_{2ij}	0,6087	0,8168	1	0,8944	0,8953	0,8929	0,8859	0,882	0,8752
Y_{3ij}	0,5541	0,8041	0,8944	1	0,9337	0,9368	0,9352	0,9337	0,9308
Y_{4ij}	0,5162	0,7898	0,8953	0,9337	1	0,9561	0,9578	0,9586	0,9571
Y_{5ij}	0,4899	0,7758	0,8929	0,9368	0,9561	1	0,9681	0,9706	0,9704
Y_{6ij}	0,464	0,7628	0,8859	0,9352	0,9578	0,9681	1	0,9771	0,9781
Y_{7ij}	0,4502	0,7547	0,882	0,9337	0,9586	0,9706	0,9771	1	0,9815
Y_{8ij}	0,4334	0,7435	0,8752	0,9308	0,9571	0,9704	0,9781	0,9815	1

COMBINACIÓN 15

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,5357**	**0,3204**	**0,1727**	**0,0622**	**-0,0004**	**-0,0484**	**-0,0816**	**-0,1052**
Y_{1ij}	0,5357	1	**0,6403**	**0,5784**	**0,5137**	**0,4678**	**0,4335**	**0,41**	**0,3922**
Y_{2ij}	0,3204	0,6403	1	**0,7979**	**0,7852**	**0,7656**	**0,7501**	**0,7356**	**0,7241**
Y_{3ij}	0,1727	0,5784	0,7979	1	0,8839	0,8817	0,8759	0,8694	0,8629
Y_{4ij}	0,0622	0,5137	0,7852	0,8839	1	0,9278	0,9297	0,9291	0,9261
Y_{5ij}	-0,0004	0,4678	0,7656	0,8817	0,9278	1	0,9511	0,9541	0,954
Y_{6ij}	-0,0484	0,4335	0,7501	0,8759	0,9297	0,9511	1	0,966	0,9673
Y_{7ij}	-0,0816	0,41	0,7356	0,8694	0,9291	0,9541	0,966	1	0,9748
Y_{8ij}	-0,1052	0,3922	0,7241	0,8629	0,9261	0,954	0,9673	0,9748	1

COMBINACIÓN 16

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,4429**	**0,0539**	**-0,2169**	**-0,3508**	**-0,4248**	**-0,4676**	**-0,4969**	**-0,5157**
Y_{1ij}	0,4429	1	**0,3668**	**0,2118**	**0,126**	**0,0603**	**0,0262**	**0,0002**	**-0,0174**
Y_{2ij}	0,0539	0,3668	1	**0,6229**	**0,6056**	**0,5824**	**0,5651**	**0,5503**	**0,5409**
Y_{3ij}	-0,2169	0,2118	0,6229	1	0,8134	0,816	0,816	0,8104	0,8073
Y_{4ij}	-0,3508	0,126	0,6056	0,8134	1	0,8969	0,9043	0,9048	0,904
Y_{5ij}	-0,4248	0,0603	0,5824	0,816	0,8969	1	0,9354	0,9391	0,9417
Y_{6ij}	-0,4676	0,0262	0,5651	0,816	0,9043	0,9354	1	0,9576	0,9606
Y_{7ij}	-0,4969	0,0002	0,5503	0,8104	0,9048	0,9391	0,9576	1	0,9688
Y_{8ij}	-0,5157	-0,0174	0,5409	0,8073	0,904	0,9417	0,9606	0,9688	1

Como en todos los modelos anteriores de la subsituación 17, se sigue manteniendo el patrón decreciente entre la primera ocasión de medida y el resto de ocasiones.

El modelo de la combinación 16, muestra los valores de correlación más bajos, en comparación con el resto de modelos de esta subsituación, especialmente, cuando el número de mediciones no es elevado. Esto se debe a que se fijó un valor de correlación entre el intercepto y la pendiente, tanto en el nivel de aula como de estudiante, alto, pero de direccionalidad negativa (-0,9). En este modelo para obtener estimaciones fiables es necesario realizar más de seis mediciones.

En el modelo de la combinación 13, se fijan los mismos valores de correlación entre el intercepto y la pendiente, que los fijados para la combinación 4; solo que en este último modelo se fijó a nivel de aula una correlación de 0,2 y a nivel de estudiante de -0,9 y, en la combinación 13 es al revés. Es en el modelo de la combinación 13, donde los valores de correlación de las puntuaciones de los estudiantes en las distintas ocasiones de medida, son más fiables. Por ejemplo, en dicho modelo la correlación entre la ocasión dos y tres es de 0,725, mientras que en el modelo de la combinación 4, es de 0,583.

En la combinación 14, se fijaron los mismos valores de correlación entre el intercepto y la pendiente que en la combinación 8, ya que, en ambos casos la correlación se fijó a -0,9 y 0,9, solo que en la combinación 14 el 0,9 se fijó en la correlación entre el intercepto y la pendiente del nivel de estudiante y en el modelo de la combinación 8, en el nivel de aula. Este último modelo muestra correlaciones inferiores, así, por ejemplo, mientras que en dicho modelo la correlación entre la ocasión dos y tres es de 0,649, en el modelo de la combinación 14, es de 0,817.

Finalmente, en la combinación 15, se fijaron los mismos valores de correlación que en la combinación 12, solo que en la 12, a nivel de aula, la correlación entre el intercepto y la pendiente se fijó a -0,2 y en el nivel de estudiante a -0,9. En la combinación 15, estos valores se encuentran invertidos, proporcionando estimaciones mas fiables con un menor número de ocasiones de medida, ya que, por ejemplo, la correlación entre la ocasión dos y tres es de 0,640 y, en el modelo de la combinación 12, era de 0,496. Pero en ambos modelos con cinco o seis mediciones, las correlaciones son superiores a 0,90.

A la vista de los resultados de los distintos modelos de la subsituación 17, se observa que el patrón de correlación entre la primera ocasión de medida y el resto de ocasiones tiende a ser decreciente, aunque se fijen valores de correlación entre el intercepto y la pendiente en alguno de los niveles (aula o estudiante) o en ambos niveles (aula y estudiante), de intensidad alta y direccionalidad positiva (0,9). Por lo tanto, en estos modelos y siguiendo a Nesselroade et al., (1980), el ERM se incrementa a medida que se añaden mediciones. Además, dicho patrón simplex se extiende en todos los modelos mas allá, de la segunda ocasión de medida.

Asímismo, se observa que cuando en dos modelos, se fijan los mismos valores de correlación entre el intercepto y la pendiente, sólo que en uno de ellos es mayor el valor fijado en el nivel de aula y en otro el fijado en el nivel de estudiante, se obtienen correlaciones bastante similares. Aunque en determinados modelos, se observa una tendencia, donde generalmente, los valores de correlación son ligeramente superiores en el modelo donde la correlación mayor se ha fijado en el nivel de estudiante. Esto puede ser debido, a que, la muestra en este nivel esta compuesta por 10.000 unidades y en el nivel de aula por 400. Dicha diferencia entre ambos modelos desaparece cuando el número de mediciones son elevadas.

Esta divergencia, entre dos modelos en los cuales se fijan los mismos valores de correlación entre el intercepto y la pendiente, es más marcada cuando dichos valores son muy distintos. Muestra de ello, son, por ejemplo, los resultados obtenidos en los modelos de la combinación 10 y 7, donde la correlación entre el intercepto y la pendiente del nivel dos y tres se fijó a -0,2 y 0,9, respectivamente. En estos modelos, la precisión con la que se estima las puntuaciones de los estudiantes difieren mas entre sí, que en los modelos de las combinaciones 9 y 3, donde la correlación entre el intercepto y la pendiente del nivel de aula era de -0,2 y la del nivel de estudiante de 0,2.

De todos los modelos presentados en la subsituación 17, el modelo que presenta estimaciones más fiables con menos ocasiones de medida, es el modelo de la combinación 6, ya que, en el mismo, se considera que, tanto a nivel de aula, como a nivel de estudiante, el rendimiento de partida esta muy relacionado con el que se alcanza a lo largo del tiempo. Si los niveles de logro iniciales son altos, también lo son los obtenidos a lo largo de las mediciones, pero si, por el contrario, son bajos la pendiente de crecimiento no se incrementa mucho (correlación entre el intercepto y

la pendiente de 0,9). En este modelo con 3 mediciones se obtienen estimaciones de las puntuaciones de los estudiantes muy fiables (correlación superior a 0,90).

Por el contrario, el modelo que presenta estimaciones menos fiables con el mismo número de mediciones que el resto de modelos, es el correspondiente a la combinación 16. Esto es debido a que se fijarón unos valores de correlación entre el intercepto y la pendiente del nivel de estudiante y del nivel de aula, de intensidad alta, pero con direccionalidad negativa (-0,9). En este modelo, se requieren unas cinco o seis mediciones, para poder predecir las puntuaciones de los estudiantes en los cursos posteriores con cierta precisión (valores de correlación superior a 0,90).

En los resultados de los modelos de la **subsituación 18** (Tabla VI.44, Tabla VI.45, Tabla VI.46 y Tabla VI.47), que se presentan a continuación, la varianza de los coeficientes aleatorios se fijó a un 1,25%, a excepción de la varianza del error intrasujeto, que se fijó a un 95%. Los modelos de esta subsituación, son similares al modelo 7 y a los modelos de la situación 10, ya que, en todos ellos la varianza de la pendiente y del intercepto suman un 2,5%, y la varianza del error intrasujeto un 95%. En el modelo 7 y en los modelos de la situación 10, solo se fijó varianza en la pendiente y en el estatus de uno de los niveles, mientras que en la subsituación 18, se fijó varianza en el estatus y la pendiente del nivel dos y tres.

Tabla VI.44. *Matriz de correlaciones de las combinaciones 1, 2, 3 y 4 de la subsituación 18*

SUBSITUACIÓN 18

EFECTOS FIJOS	Coef.
β_{000}	4,6
β_{100}	0,4
EFECTOS ALEATORIOS	
NIVEL 3	$\sigma(u_{00j})$ $\sigma(u_{10j})$
$\sigma(u_{00j})$	1,25%
$\sigma(u_{10j})$	1,25%
NIVEL 2	$\sigma(r_{0ij})$ $\sigma(r_{1ij})$
$\sigma(r_{0ij})$	1,25%
$\sigma(r_{1ij})$	1,25%
NIVEL 1	$\sigma(e_{tij})$
$\sigma(e_{tij})$	95%

Raulas y Restudiantes

COMBINACIÓN 1

R_{aulas}: 0,2

$R_{estudiantes}$:0,2

COMBINACIÓN 2

R_{aulas}: 0,2

$R_{estudiantes}$:0,9

COMBINACIÓN 3

R_{aulas}: 0,2

$R_{estudiantes}$: -0,2

COMBINACIÓN 4

R_{aulas}: 0,2

$R_{estudiantes}$: -0,9

MATRIZ DE CORRELACIONES

COMBINACIÓN 1

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,0229**	**0,0378**	**0,053**	**0,0287**	**0,0269**	**0,023**	**0,0284**	**0,0287**
Y_{1ij}	0,0229	1	**0,0968**	0,0998	0,1376	0,1511	0,1426	0,1525	0,1676
Y_{2ij}	0,0378	0,0968	1	0,1771	0,2094	0,2311	0,2481	0,2638	0,2887
Y_{3ij}	0,053	0,0998	0,1771	1	0,2838	0,3196	0,3386	0,3729	0,3935
Y_{4ij}	0,0287	0,1376	0,2094	0,2838	1	0,3845	0,4129	0,4429	0,4710
Y_{5ij}	0,0269	0,1511	0,2311	0,3196	0,3845	1	0,4711	0,5135	0,5348
Y_{6ij}	0,023	0,1426	0,2481	0,3386	0,4129	0,4711	1	0,5453	0,5686
Y_{7ij}	0,0284	0,1525	0,2638	0,3729	0,4429	0,5135	0,5453	1	**0,6151**
Y_{8ij}	0,0287	0,1676	0,2887	0,3935	0,471	0,5348	0,5686	0,6151	1

COMBINACIÓN 2

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,03**	**0,0741**	**0,0548**	**0,0771**	**0,0763**	**0,0696**	**0,0796**	**0,0851**
Y_{1ij}	0,03	1	**0,1094**	0,126	0,1569	0,163	0,1789	0,1923	0,2079
Y_{2ij}	0,0741	0,1094	1	0,219	0,2459	0,2924	0,2974	0,319	0,3396
Y_{3ij}	0,0548	0,126	0,219	1	0,3095	0,3651	0,3852	0,4032	0,4258
Y_{4ij}	0,0771	0,1569	0,2459	0,3095	1	0,4027	0,4397	0,4649	0,4919
Y_{5ij}	0,0763	0,163	0,2924	0,3651	0,4027	1	0,5014	0,5375	0,5579
Y_{6ij}	0,0696	0,1789	0,2974	0,3852	0,4397	0,5014	1	0,5746	0,6063
Y_{7ij}	0,0796	0,1923	0,319	0,4032	0,4649	0,5375	0,5746	1	**0,646**
Y_{8ij}	0,0851	0,2079	0,3396	0,4258	0,4919	0,5579	0,6063	0,646	1

COMBINACIÓN 3

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,0421**	**0,0352**	**0,0388**	**0,0243**	**0,0175**	**0,0258**	**0,0221**	**0,0196**
Y_{1ij}	0,0421	1	**0,0761**	0,0644	0,0884	0,0963	0,1107	0,1221	0,1298
Y_{2ij}	0,0352	0,0761	1	0,1447	0,1768	0,2013	0,2121	0,2262	0,2362
Y_{3ij}	0,0388	0,0644	0,1447	1	0,2506	0,2849	0,3223	0,3379	0,3495
Y_{4ij}	0,0243	0,0884	0,1768	0,2506	1	0,3288	0,3761	0,4024	0,4135
Y_{5ij}	0,0175	0,0963	0,2013	0,2849	0,3288	1	0,4284	0,475	0,4827
Y_{6ij}	0,0258	0,1107	0,2121	0,3223	0,3761	0,4284	1	0,5256	0,5495
Y_{7ij}	0,0221	0,1221	0,2262	0,3379	0,4024	0,475	0,5256	1	**0,5872**
Y_{8ij}	0,0196	0,1298	0,2362	0,3495	0,4135	0,4827	0,5495	0,5872	1

COMBINACIÓN 4

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,0172**	**-0,0029**	**0,0093**	**-0,0041**	**-0,0334**	**-0,0172**	**-0,021**	**-0,026**
Y_{1ij}	0,0172	1	**0,0693**	0,0609	0,0846	0,0752	0,095	0,1048	0,0927
Y_{2ij}	-0,0029	0,0693	1	0,1543	0,1562	0,1686	0,2161	0,2124	0,215
Y_{3ij}	0,0093	0,0609	0,1543	1	0,2236	0,2585	0,2869	0,3069	0,3255
Y_{4ij}	-0,0041	0,0846	0,1562	0,2236	1	0,3217	0,3592	0,3848	0,4052
Y_{5ij}	-0,0334	0,0752	0,1686	0,2585	0,3217	1	0,4254	0,4534	0,4841
Y_{6ij}	-0,0172	0,095	0,2161	0,2869	0,3592	0,4254	1	0,5064	0,5417
Y_{7ij}	-0,021	0,1048	0,2124	0,3069	0,3848	0,4534	0,5064	1	**0,5768**
Y_{8ij}	-0,026	0,0927	0,215	0,3255	0,4052	0,4841	0,5417	0,5768	1

En estos modelos de la combinación 1 a la 4, donde el porcentaje fijado en la varianza del error intrasujeto es muy elevado en comparación con el fijado en el resto de coeficientes aleatorios, el patrón de correlación entre la primera ocasión de medida y el resto de ocasiones tiende a ser constante. Esto también ocurría en el modelo 7, en el cual, la correlación entre el intercepto y la pendiente en ambos niveles se fijó a 0, y en los modelos de la situación 10, donde se fijó la correlación entre el intercepto y la pendiente del nivel de estudiante a 0,2, -0,2, -0,9. En estos modelos de la subsituación 10, cuando dicha correlación se fijaba a 0,9 (modelo 10.1), ya se observaba una cierta tendencia hacia un patrón creciente. Sin embargo, en la combinación 2 de la subsituación 17, donde la correlación entre la pendiente y el estatus de nivel del estudiante se fijó a 0,9, se observa un patrón constante y nulo, en torno a 0.

Si se comparan los resultados del modelo de la combinación 2, con el modelo 10.1, ambos presentan valores de correlación muy similares. Pues, por ejemplo, en el modelo 10.1, la correlación entre la ocasión dos y tres era de 0,137 y, en el modelo de la combinación 2, es de 0,109. Con nueve ocasiones de medida, la correlación entre la ocasión ocho y nueve en el modelo de la combinación 2 es de 0,646 y en el modelo 10.1 era de 0,643. Esta semejanza entre ambos modelos, puede ser debido a que, en los mismos se fijó una correlación entre el intercepto y la pendiente en el nivel de estudiante de 0,9 y, aunque en el modelo de la combinación 2, se fijó un valor de correlación entre el estatus y la pendiente a nivel de aula (0,2), cosa que no ocurre en el modelo 10.1, como dicho valor de correlación es bajo no afecta mucho en la fiabilidad de las estimaciones de las puntuaciones de los estudiantes.

Si se compara el modelo de la combinación 2, con el resto de modelos presentados en la tabla anterior (Tabla VI.44), se observa que es el modelo que presenta los valores de correlación más elevados. Aunque, ni con nueve ocasiones de medida, las estimaciones de las puntuaciones de los estudiantes, son muy precisas, ya que, las correlaciones entre la ocasión ocho y nueve son inferiores a 0,70.

Al comparar los resultados del modelo 10, donde se fijó una correlación entre el intercepto y la pendiente del nivel de estudiante de 0,2, con los del modelo de la combinación 1, donde además de fijar dicha correlación a nivel de estudiante, se fijó a nivel de aula una correlación de 0,2, se obtiene en ambos modelos valores de correlación muy similares. Así, por ejemplo, en el modelo de la combinación 1, la

correlación entre la ocasión dos y tres es de 0,097, mientras que en el modelo 10, era de 0,1. Con nueve ocasiones de medida, en el modelo 10, la correlación era de 0,610 y en el modelo de la combinación 1 es de 0,646.

En la combinación 3, en la cual se fijó una correlación entre el estatus y la pendiente del nivel de estudiante de -0,2, muestra, también, valores de correlación muy semejantes al modelo 10.2, en el cual la correlación entre el intercepto y la pendiente del nivel de estudiante también era de -0,2. Sin embargo, el modelo de la combinación 4, difiere algo más del modelo 10.3, a pesar de que en ambos modelos la correlación del nivel del estudiante se fijó a -0,9. En el modelo de la combinación 4, los valores son ligeramente superiores, así, por ejemplo, en este modelo la correlación entre la tercera y cuarta ocasión es de 0,154 y en el modelo 10.3 era de 0,058, aunque, con nueve ocasiones de medida la correlación en el modelo 10.3 era de 0,522 y en el modelo de la combinación 4 es de 0,576.

En la siguiente tabla se muestra los resultados para las **combinaciones de la 5 a la 8.**

Tabla VI.45. *Matriz de correlaciones de las combinaciones 5, 6, 7 y 8 de la subsituación 18*

SUBSITUACIÓN 18

EFECTOS FIJOS	Coef.
β_{000}	4,6
β_{100}	0,4
EFECTOS ALEATORIOS	
NIVEL 3	$\sigma(u_{00j})$ $\sigma(u_{10j})$
$\sigma(u_{00j})$	1,25%
$\sigma(u_{10j})$	1,25%
NIVEL 2	$\sigma(r_{0ij})$ $\sigma(r_{1ij})$
$\sigma(r_{0ij})$	1,25%
$\sigma(r_{1ij})$	1,25%
NIVEL 1	$\sigma(e_{tij})$
$\sigma(e_{tij})$	95%

Raulas y Restudiantes

COMBINACIÓN 5

R_{aulas}: 0,9

$R_{estudiantes}$:0,2

COMBINACIÓN 6

R_{aulas}: 0,9

$R_{estudiantes}$:0,9

COMBINACIÓN 7

R_{aulas}: 0,9

$R_{estudiantes}$: -0,2

COMBINACIÓN 8

R_{aulas}: 0,9

$R_{estudiantes}$: -0,9

MATRIZ DE CORRELACIONES

COMBINACIÓN 5

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	0,0421	0,0438	0,0508	0,0634	0,0762	0,0748	0,0661	0,0716
Y_{1ij}	0,0421	1	**0,1047**	0,1398	0,158	0,1857	0,1893	0,2114	0,2161
Y_{2ij}	0,0438	0,1047	1	0,1991	0,2292	0,2681	0,2879	0,3092	0,3118
Y_{3ij}	0,0508	0,1398	0,1991	1	0,2902	0,3241	0,3578	0,3896	0,3922
Y_{4ij}	0,0634	0,158	0,2292	0,2902	1	0,3967	0,4301	0,4609	0,4784
Y_{5ij}	0,0762	0,1857	0,2681	0,3241	0,3967	1	0,4902	0,5157	0,5303
Y_{6ij}	0,0748	0,1893	0,2879	0,3578	0,4301	0,4902	1	0,5704	0,5939
Y_{7ij}	0,0661	0,2114	0,3092	0,3896	0,4609	0,5157	0,5704	1	**0,6252**
Y_{8ij}	0,0716	0,2161	0,3118	0,3922	0,4784	0,5303	0,5939	0,6252	1

COMBINACIÓN 6

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,052**	**0,0581**	**0,0829**	**0,0867**	**0,1092**	**0,1107**	**0,1202**	**0,1232**
Y_{1ij}	0,052	1	**0,1104**	0,1445	0,1834	0,1992	0,2213	0,2395	0,2444
Y_{2ij}	0,0581	0,1104	1	0,2082	0,2487	0,2813	0,287	0,3122	0,3258
Y_{3ij}	0,0829	0,1445	0,2082	1	0,3106	0,3389	0,3707	0,3937	0,4065
Y_{4ij}	0,0867	0,1834	0,2487	0,3106	1	0,4165	0,446	0,4781	0,5063
Y_{5ij}	0,1092	0,1992	0,2813	0,3389	0,4165	1	0,4992	0,5335	0,5582
Y_{6ij}	0,1107	0,2213	0,287	0,3707	0,446	0,4992	1	0,5776	0,6095
Y_{7ij}	0,1202	0,2395	0,3122	0,3937	0,4781	0,5335	0,5776	1	**0,6429**
Y_{8ij}	0,1232	0,2444	0,3258	0,4065	0,5063	0,5582	0,6095	0,6429	1

COMBINACIÓN 7

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	0,0361	0,0397	0,0481	0,0461	0,0621	0,0679	0,0633	0,0639
Y_{1ij}	0,0361	1	**0,0763**	0,1132	0,1257	0,1317	0,1558	0,17	0,1694
Y_{2ij}	0,0397	0,0763	1	0,1846	0,2219	0,2469	0,2615	0,2805	0,302
Y_{3ij}	0,0481	0,1132	0,1846	1	0,3067	0,3265	0,3821	0,383	0,4035
Y_{4ij}	0,0461	0,1257	0,2219	0,3067	1	0,4003	0,4339	0,4557	0,4741
Y_{5ij}	0,0621	0,1317	0,2469	0,3265	0,4003	1	0,4878	0,5107	0,5362
Y_{6ij}	0,0679	0,1558	0,2615	0,3821	0,4339	0,4878	1	0,5662	0,5946
Y_{7ij}	0,0633	0,17	0,2805	0,383	0,4557	0,5107	0,5662	1	0,625
Y_{8ij}	0,0639	0,1694	0,302	0,4035	0,4741	0,5362	0,5946	0,625	1

COMBINACIÓN 8

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	0,0253	0,0382	0,0267	0,0492	0,0383	0,0246	0,0307	0,0292
Y_{1ij}	0,0253	1	**0,0913**	0,0862	0,0893	0,1149	0,1263	0,143	0,1305
Y_{2ij}	0,0382	0,0913	1	0,1532	0,2039	0,2264	0,2509	0,2576	0,2678
Y_{3ij}	0,0267	0,0862	0,1532	1	0,2624	0,3055	0,3367	0,3689	0,3825
Y_{4ij}	0,0492	0,0893	0,2039	0,2624	1	0,3643	0,403	0,4405	0,4637
Y_{5ij}	0,0383	0,1149	0,2264	0,3055	0,3643	1	0,4587	0,498	0,5142
Y_{6ij}	0,0246	0,1263	0,2509	0,3367	0,403	0,4587	1	0,5481	0,5787
Y_{7ij}	0,0307	0,143	0,2576	0,3689	0,4405	0,498	0,5481	1	0,6127
Y_{8ij}	0,0292	0,1305	0,2678	0,3825	0,4637	0,5142	0,5787	0,6127	1

En los modelos anteriores, se sigue manteniendo el patrón constante entre la primera ocasión de medida y el resto de ocasiones, a excepción del modelo de la combinación 6, en el cual se observa, una cierta tendencia hacia un patrón creciente. Este modelo es el que presenta los valores de correlación mas elevados, aunque, con nueve mediciones, las estimaciones de las puntuaciones de los sujetos, siguen siendo poco fiables, ya que, las correlaciones son inferiores a 0,70.

Nuevamente los modelos de la tabla anterior VI.45, guardan ciertos parecidos con los modelos de la situación 10, ya que, las correlaciones fijadas entre el intercepto y la pendiente del nivel de estudiante son coincidentes. Así, por ejemplo, en el modelo de la combinación 5, en el cual se fijó el mismo valor de correlación entre el inntercepto y la pendiente del nivel de estudiante, que en el modelo 10, se obtienen resultados muy parecidos, a pesar de que en el modelo de la combinación 5, la correlación entre el intercepto y la pendiente del nivel de aula se fijó a 0,9 y en el modelo 10 a 0, puesto que en este último modelo no se estableció varianza ni en el intercepto ni en la pendiente del nivel de aula. Así, en el modelo de la combinación 5, la correlación entre la ocasión dos y tres es de 0,105 mientras que en el modelo 10 era de 0,100; con nueve ocasiones de medida en el modelo 10, la correlación era de 0,610 y, en la combinación 5 es de 0,625.

En este modelo de la combinación 5 se fijan los mismos valores de correlación entre el intercepto y la pendiente que en el modelo de la combinación 2. Si se comparan los resultados de ambos modelos son similares, por ejemplo, mientras que en el modelo de la combinación 2 la correlación entre la ocasión dos y tres era de 0,194 en el modelo de la combinación 5 es de 0,104.

En el modelo de la combinación 6, se fijó una correlación entre el intercepto y la pendiente del nivel de aula y del nivel de estudiante de 0,9. Este mismo valor de correlación, para el nivel de estudiante también se fijó en el modelo 10.1, que muestra correlaciones muy similares a las del modelo de la combinación 6, ya que, por ejemplo, la correlación entre la ocasión dos y tres en el modelo 10.1, era de 0,137 y, en el modelo de la combinación 6 es de 0,110.

Diferencias más marcadas existen entre los resultados del modelo de la combinación 7 y el modelo 10.2, en los cuales, la correlación entre el intercepto y la pendiente en el nivel de estudiante se fijó a -0,2. El modelo 10.2, muestra valores de correlación ligeramente inferiores a los de la combinación 7, ya que, por ejemplo, la correlación entre la ocasión dos y tres de este modelo es de 0,076, mientras que, en el modelo 10.3, era de 0,051.

Finalmente, la combinación 8 también presenta correlaciones ligeramente superiores a las del modelo 10.3, en el cual la correlación entre el intercepto y la pendiente del nivel de estudiante también se fijó a -0,9, pero no existía relación entre el intercepto y la pendiente del nivel de aula. En la combinación 8, la correlación entre la ocasión dos y tres es de 0,091. mientras que, en el modelo 10.3 era de 0,013

En la tabla siguiente VI.46, dentro de la subsituación 18, se muestran los resultados de las **combinaciones 9, 10, 11 y 12.**

Tabla VI.46. *Matriz de correlaciones de las combinaciones 9, 10, 11 y12 de la subsituación 18*

SUBSITUACIÓN 18

EFECTOS FIJOS	Coef.
β_{000}	4,6
β_{100}	0,4
EFECTOS ALEATORIOS	
NIVEL 3	$\sigma(u_{00j})$ $\sigma(u_{10j})$
$\sigma(u_{00j})$	1,25%
$\sigma(u_{10j})$	1,25%
NIVEL 2	$\sigma(r_{0ij})$ $\sigma(r_{10j})$
$\sigma(r_{0ij})$	1,25%
$\sigma(r_{10j})$	1,25%
NIVEL 1	$\sigma(e_{tij})$
$\sigma(e_{tij})$	95%

Raulas y Restudiantes

COMBINACIÓN 9

R_{aulas}: -0,2

$R_{estudiantes}$:0,2

COMBINACIÓN 10

R_{aulas}: -0,2

$R_{estudiantes}$:0,9

COMBINACIÓN 11

R_{aulas}: -0,2

$R_{estudiantes}$: -0,2

COMBINACIÓN 12

R_{aulas}: -0,2

$R_{estudiantes}$: -0,9

MATRIZ DE CORRELACIONES

COMBINACIÓN 9

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	0,0207	0,0121	0,0155	0,0274	0,0086	0,0104	0,0164	-0,0056
Y_{1ij}	0,0207	1	0,062	0,0804	0,104	0,1062	0,116	0,1255	0,1242
Y_{2ij}	0,0121	0,062	1	0,1513	0,1799	0,1905	0,2193	0,2297	0,2439
Y_{3ij}	0,0155	0,0804	0,1513	1	0,2515	0,2847	0,3185	0,3476	0,3568
Y_{4ij}	0,0274	0,104	0,1799	0,2515	1	0,3392	0,3727	0,4116	0,4283
Y_{5ij}	0,0086	0,1062	0,1905	0,2847	0,3392	1	0,449	0,467	0,4973
Y_{6ij}	0,0104	0,116	0,2193	0,3185	0,3727	0,449	1	0,5221	0,5544
Y_{7ij}	0,0164	0,1255	0,2297	0,3476	0,4116	0,467	0,5221	1	**0,5848**
Y_{8ij}	-0,0056	0,1242	0,2439	0,3568	0,4283	0,4973	0,5544	0,5848	1

COMBINACIÓN 10

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	0,0532	0,0459	0,0252	0,0399	0,0503	0,0445	0,0396	0,0472
Y_{1ij}	0,0532	1	**0,0968**	0,1112	0,1287	0,1438	0,1714	0,1719	0,1774
Y_{2ij}	0,0459	0,0968	1	0,1921	0,2213	0,2606	0,2738	0,302	0,308
Y_{3ij}	0,0252	0,1112	0,1921	1	0,2951	0,3432	0,37	0,3959	0,4005
Y_{4ij}	0,0399	0,1287	0,2213	0,2951	1	0,3948	0,4199	0,4608	0,4657
Y_{5ij}	0,0503	0,1438	0,2606	0,3432	0,3948	1	0,4825	0,5231	0,5431
Y_{6ij}	0,0445	0,1714	0,2738	0,37	0,4199	0,4825	1	0,5575	0,5831
Y_{7ij}	0,0396	0,1719	0,302	0,3959	0,4608	0,5231	0,5575	1	0,6269
Y_{8ij}	0,0472	0,1774	0,308	0,4005	0,4657	0,5431	0,5831	0,6269	1

COMBINACIÓN 11

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	0,043	0,0161	0,0188	0,008	-0,0011	-0,0101	0,0103	-0,0084
Y_{1ij}	0,043	1	**0,0434**	0,0623	0,0943	0,102	0,1015	0,1115	0,115
Y_{2ij}	0,0161	0,0434	1	0,1469	0,1615	0,1955	0,1948	0,2164	0,2205
Y_{3ij}	0,0188	0,0623	0,1469	1	0,24	0,2664	0,2898	0,3098	0,3248
Y_{4ij}	0,008	0,0943	0,1615	0,24	1	0,3333	0,3828	0,4013	0,4272
Y_{5ij}	-0,0011	0,102	0,1955	0,2664	0,3333	1	0,4288	0,4732	0,4888
Y_{6ij}	-0,0101	0,1015	0,1948	0,2898	0,3828	0,4288	1	0,5054	0,5354
Y_{7ij}	0,0103	0,1115	0,2164	0,3098	0,4013	0,4732	0,5054	1	0,5845
Y_{8ij}	-0,0084	0,115	0,2205	0,3248	0,4272	0,4888	0,5354	0,5845	1

COMBINACIÓN 12

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	0,0121	-0,0145	-0,0016	-0,0103	-0,0427	-0,0529	-0,0496	-0,0623
Y_{1ij}	0,0121	1	0,0446	0,0577	0,0707	0,0764	0,0612	0,0711	0,0819
Y_{2ij}	-0,0145	0,0446	1	0,1192	0,1332	0,1575	0,1829	0,187	0,2
Y_{3ij}	-0,0016	0,0577	0,1192	1	0,1828	0,2289	0,2569	0,2783	0,2915
Y_{4ij}	-0,0103	0,0707	0,1332	0,1828	1	0,2897	0,3302	0,3561	0,3738
Y_{5ij}	-0,0427	0,0764	0,1575	0,2289	0,2897	1	0,3921	0,4323	0,4594
Y_{6ij}	-0,0529	0,0612	0,1829	0,2569	0,3302	0,3921	1	0,4936	0,5125
Y_{7ij}	-0,0496	0,0711	0,187	0,2783	0,3561	0,4323	0,4936	1	0,5694
Y_{8ij}	-0,0623	0,0819	0,2	0,2915	0,3738	0,4594	0,5125	0,5694	1

En los modelos de la tabla anterior VI.46, también entre la primera ocasión de medida y el resto de ocasiones, se mantiene el patrón constante que viene predominando en la mayoría de los modelos de esta subsituación 18.

Si se comparan los resultados de la combinación 9, con los resultados del modelo 10, se obtienen correlaciones muy semejantes, aunque en determinadas ocasiones son ligeramente superiores en el modelo 10. Esto puede deberse a que en el modelo 10, la correlación entre el intercepto y la pendiente del nivel de aula se fijó a 0 y en el modelo de la combinación 9 se fijó a -0,2. Así, por ejemplo, mientras que en el modelo de la combinación 9, la correlación entre la ocasión ocho y nueve es de 0,585, en el modelo 10 era de 0,6103.

En el modelo de la combinación 9, la correlación entre el estatus y la pendiente del nivel de aula se fijó a -0,2 y la del nivel de estudiante a 0,2. Estos valores se fijan invertidos en el modelo de la combinación 3. En ambos modelos, las correlaciones son muy similares, pues en los mismos, con tres ocasiones de medida las correlaciones son superiores a 0,15 y con cuatro a 0,25, siendo ligeramente superiores en el modelo de la combinación 9.

En la combinación 10, se fijó la misma correlación entre el intercepto y la pendiente del nivel de estudiante que en el modelo 10.1 (0,9), por ello en ambos modelos las correlaciones son muy similares, aunque ligeramente superiores en el modelo 10.1, debido quizás a que en el mismo no se fijó relación entre el intercepto y la pendiente del nivel de aula y, en el modelo de la combinación 10 se fijó una correlación de -0,2. En este modelo, por ejemplo, la correlación entre la ocasión dos y tres es de 0,097 mientras que en el modelo 10.1 era de 0,137.

En dicha combinación 10, se fijaron los mismos valores de correlación entre el intercepto y la pendiente del nivel de aula y estudiante que en la combinación 7, pero invertidos. Este último modelo muestra valores de correlación ligeramente inferiores, ya que, por ejemplo, en dicho modelo la correlación entre la ocasión dos y tres era de 0,076 y en la combinación 10, es de 0,097.

En el modelo 10.2 se fijó una correlación entre el intercepto y la pendiente del nivel de estudiante de -0,2, como en el modelo de la combinación 11. En ambos modelos la precisión con la que se estima las puntuaciones de los estudiantes es muy similar, ya que, por ejemplo, la correlación entre la ocasión tres y cuatro es de 0,147 y en el modelo 10.2 era de 0,124.

Finalmente, el modelo de la combinación 12, tiene en común con el modelo 10.3, en que, en ambos se fijó la misma correlación entre el intercepto y la pendiente del nivel de estudiante (-0,9). Pero a nivel de aula, la correlación entre el intercepto y la pendiente en el modelo 10.3 se fijó a 0 y en el modelo de la combinación 12 se fijó a -0,2. Los dos modelos en general cuentan con correlaciones que son muy semejantes, en el modelo 10.3, la correlación entre la ocasión cuatro y cinco era de 0,143, y en el modelo de la combinación 12 es de 0,182, aunque en algunas ocasiones este último modelo muestra valores ligeramente superiores, por ejemplo, la correlación entre la ocasión tres y cuatro es de 0,119 y en el modelo 10.3, era de 0,058.

Posteriormente, se muestran los modelos de las **combinaciones 13 a la 16.**

Tabla VI.47. *Matriz de correlaciones de las combinaciones 13, 14, 15 y16 de la subsituación 18*

SUBSITUACIÓN 18

EFECTOS FIJOS	Coef.
β_{000}	**4,6**
β_{100}	**0,4**
EFECTOS ALEATORIOS	
NIVEL 3	$\sigma(u_{00j})$ $\sigma(u_{10j})$
$\sigma(u_{00j})$	**1,25%**
$\sigma(u_{10j})$	**1,25%**
NIVEL 2	$\sigma(r_{0ij})$ $\sigma(r_{1ij})$
$\sigma(r_{0ij})$	**1,25%**
$\sigma(r_{1ij})$	**1,25%**
NIVEL 1	$\sigma(e_{tij})$
$\sigma(e_{tij})$	**95%**

Raulas y Restudiantes

COMBINACIÓN 13

R_{aulas}: **-0,9**

$R_{estudiantes}$:**0,2**

COMBINACIÓN 14

R_{aulas}: **-0,9**

$R_{estudiantes}$:**0,9**

COMBINACIÓN 15

R_{aulas}: **-0,9**

$R_{estudiantes}$: **-0,2**

COMBINACIÓN 16

R_{aulas}: **-0,9**

$R_{estudiantes}$: **-0,9**

MATRIZ DE CORRELACIONES

COMBINACIÓN 13

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	0,0211	0,0209	0,0036	-0,0195	-0,0051	-0,0263	-0,0348	-0,0388
Y_{1ij}	0,0211	1	0,0652	0,0752	0,0776	0,0964	0,0984	0,0917	0,1135
Y_{2ij}	0,0209	0,0652	1	0,1159	0,1477	0,1569	0,1829	0,1871	0,2057
Y_{3ij}	0,0036	0,0752	0,1159	1	0,23	0,2622	0,2892	0,3165	0,3385
Y_{4ij}	-0,0195	0,0776	0,1477	0,23	1	0,3351	0,3586	0,3919	0,4145
Y_{5ij}	-0,0051	0,0964	0,1569	0,2622	0,3351	1	0,438	0,4494	0,4721
Y_{6ij}	-0,0263	0,0984	0,1829	0,2892	0,3586	0,438	1	0,512	0,5449
Y_{7ij}	-0,0348	0,0917	0,1871	0,3165	0,3919	0,4494	0,512	1	**0,5873**
Y_{8ij}	-0,0388	0,1135	0,2057	0,3385	0,4145	0,4721	0,5449	0,5873	1

COMBINACIÓN 14

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	0,0267	0,0362	0,0246	0,0298	0,0317	0,0291	0,0327	0,0249
Y_{1ij}	0,0267	1	**0,0801**	0,0949	0,1172	0,1193	0,1435	0,1522	0,1473
Y_{2ij}	0,0362	0,0801	1	0,1518	0,1849	0,2173	0,2334	0,2475	0,2462
Y_{3ij}	0,0246	0,0949	0,1518	1	0,2511	0,2902	0,3206	0,3352	0,3542
Y_{4ij}	0,0298	0,1172	0,1849	0,2511	1	0,3467	0,3856	0,4142	0,4303
Y_{5ij}	0,0317	0,1193	0,2173	0,2902	0,3467	1	0,4329	0,4744	0,4964
Y_{6ij}	0,0291	0,1435	0,2334	0,3206	0,3856	0,4329	1	0,5298	0,5565
Y_{7ij}	0,0327	0,1522	0,2475	0,3352	0,4142	0,4744	0,5298	1	0,5849
Y_{8ij}	0,0249	0,1473	0,2462	0,3542	0,4303	0,4964	0,5565	0,5849	1

COMBINACIÓN 15

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	0,0013	-0,0095	-0,0115	-0,0409	-0,03	-0,0477	-0,0599	-0,0651
Y_{1ij}	0,0013	1	**0,0277**	0,0434	0,0437	0,0434	0,0466	0,0604	0,0721
Y_{2ij}	-0,0095	0,0277	1	0,0907	0,1261	0,1376	0,1661	0,1763	0,1866
Y_{3ij}	-0,0115	0,0434	0,0907	1	0,1888	0,2153	0,2429	0,2647	0,278
Y_{4ij}	-0,0409	0,0437	0,1261	0,1888	1	0,282	0,3187	0,3565	0,3705
Y_{5ij}	-0,03	0,0434	0,1376	0,2153	0,282	1	0,3753	0,424	0,4505
Y_{6ij}	-0,0477	0,0466	0,1661	0,2429	0,3187	0,3753	1	0,4779	0,5035
Y_{7ij}	-0,0599	0,0604	0,1763	0,2647	0,3565	0,424	0,4779	1	0,5613
Y_{8ij}	-0,0651	0,0721	0,1866	0,278	0,3705	0,4505	0,5035	0,5613	1

COMBINACIÓN 16

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	-0,0156	-0,0223	-0,0325	-0,0535	-0,0681	-0,0802	-0,1034	-0,1071
Y_{1ij}	-0,0156	1	0,0128	0,0076	0,0285	0,0083	0,0224	0,0178	0,0255
Y_{2ij}	-0,0223	0,0128	1	0,0594	0,0721	0,0939	0,1125	0,1196	0,1423
Y_{3ij}	-0,0325	0,0076	0,0594	1	0,1458	0,1712	0,2123	0,2233	0,2533
Y_{4ij}	-0,0535	0,0285	0,0721	0,1458	1	0,2476	0,2946	0,3172	0,3411
Y_{5ij}	-0,0681	0,0083	0,0939	0,1712	0,2476	1	0,3483	0,3736	0,4138
Y_{6ij}	-0,0802	0,0224	0,1125	0,2123	0,2946	0,3483	1	0,4554	0,4887
Y_{7ij}	-0,1034	0,0178	0,1196	0,2233	0,3172	0,3736	0,4554	1	**0,5265**
Y_{8ij}	-0,1071	0,0255	0,1423	0,2533	0,3411	0,4138	0,4887	0,5265	1

En estos modelos, se sigue manteniendo el patrón constante entre la primera ocasión y el resto de mediciones. Si se comparan los resultados del modelo de la combinación 13 con los resultados del modelo 10, donde también a nivel de estudiante se fijó una correlación entre el intercepto y la pendiente de 0,2, se obtienen valores de correlación muy similares, aunque ligeramente superiores en este último modelo, ya que, por ejemplo, la correlación entre la ocasión ocho y nueve era de 0,610 y en el modelo de la combinación 13 es de 0,587.

Los valores de correlación de la combinación 13, se fijaron también en la combinación 4. Si se comparan los resultados de ambos modelos, en general, son muy similares, ya que, por ejemplo, en ambos casos la correlación entre la ocasión tres y cuatro oscila entre 0,20 y, entre la ocasión ocho y nueve alrededor de 0,55.

Si se compara el modelo 10.1, donde se fijó una correlación entre el intercepto y la pendiente en el nivel de aula de 0 y en el nivel de estudiante de 0,9, con el modelo de la combinación 14, se obtienen resultados ligeramente inferiores en este último modelo, ya que, por ejemplo, en el modelo de la combinación 14, la correlación entre la ocasión cuatro y cinco es de 0,251 y en el modelo 10.1 era de 0,341.

En este modelo de la combinación 14, se fijaron los mismos valores de correlación entre el intercepto y la pendiente que en el modelo de la combinación 8, pero invertidos, aun así, en ambos modelos las correlaciones son muy similares.

Al comparar los resultados del modelo 10.2, con los del modelo de la combinación 15, donde también se fijó a nivel de estudiante una correlación entre el intercepto y la pendiente de -0,2, los resultados son ligeramente inferiores en esta última combinación, muestra de ello, es que, por ejemplo, la correlación entre la ocasión dos y tres en el modelo de la combinación 14, es de 0,028 y en el modelo 10.2 era de 0,051.

En este modelo de la combinación 15, se fijan los mismos valores en las correlaciones entre el intercepto y la pendiente, pero invertidos, que en el modelo de la combinación 12, donde la correlación entre el estatus y la pendiente del nivel de estudiante se fijó a -0,9 y, en el nivel de aula a -0,2, en ambos modelos las correlaciones son muy similares.

Finalmente, los resultados de la combinación 16, son muy similares en todos los valores de correlación a los del modelo 10.3, donde se fijó una correlación entre el intercepto y la pendiente a nivel de estudiante de -0,9. En este modelo la correlación entre la ocasión ocho y nueve era de 0,522 y, en el modelo de la combinación 16 es de 0,527.

El modelo de la combinación 16, es el modelo que muestra los valores de correlación más bajos de todos los modelos de la subsituación 18, ya que, en ambos niveles (estudiante y aula) se fijan valores de correlación entre el intercepto y la pendiente de intensidad alta, pero de direccionalidad negativa. Por lo tanto, los estudiantes o aulas con niveles de logro iniciales altos, muestran un rendimiento que no crece mucho o se mantiene estable a lo largo del tiempo, pero los estudiantes o aulas con niveles previos de rendimiento bajo, van incrementando mucho su nivel de logro a lo largo del tiempo, siendo esta tendencia muy marcada en toda la muestra.

A la vista de los resultados de los modelos de la subsituación 18, cuando el porcentaje fijado en la varianza del error intrasujeto, es muy elevado en comparación con el resto de coeficientes aleatorios, el patrón de correlación entre la ocasión una y el resto de ocasiones se mantiene constante, por lo tanto, siguiendo a Nesselroade et al. (1980) el ERM permanece inmutable a lo largo de las diferentes ocasiones de medida. Este patrón, solo se ve alterado en la combinación 16, donde la correlación entre el intercepto y la pendiente del nivel de aula y del nivel de estudiante se fijó a 0,9. En este caso, el patrón entre la primera ocasión y el resto de ocasiones tiende a ser creciente, por lo que, el efecto de regresión a la media podría desaparecer, si se añaden un número suficiente de mediciones como para obtener una correlación de 1, pero al ser la correlación entre la ocasión ocho y nueve de 0,123, se necesitaría realizar muchas aplicaciones.

Al comparar los modelos de la situación 10, con los modelos de la subsituación 18, se observa que en general, la fiabilidad de las estimaciones no difiere mucho cuando, en dos modelos se fija el mismo valor de correlación entre el intercepto y la pendiente del nivel de estudiante.

Del mismo modo, cuando se comparan dos modelos en los cuales, se han fijado los mismos valores de correlación entre el intercepto y la pendiente, y solo cambia el nivel donde se fijan los mismos, parece que la matriz de correlaciones obtenida en

dichos modelos es muy similar, por lo tanto, el nivel donde se fija un determinado valor de correlación parece no afectar a la precisión con la que se estiman las puntuaciones de los estudiantes. A pesar de que, en determinados modelos se observa una tendencia, donde generalmente, los valores de correlación son ligeramente superiores en el modelo donde la correlación mayor se ha fijado en el nivel de estudiante, pero esa tendencia es poco marcada.

Tal y como se ha comentado en el apartado de *procedimiento de generación de modelos*, con el fin de reducir espacio, en los resultados de la siguiente subsituación, no se presentan los modelos de las 16 combinaciones, solo aquellas que se consideran que pueden tener un mayor interés. Concretamente se muestran los resultados de la combinación 1, 2, 6, 9, 10, 11, 13, 14, 15 y 16. Las demás combinaciones, no se presentan en este apartado porque los resultados no difieren mucho de los que se obtienen en las combinaciones anteriores, pero si el lector lo desea puede consultar los resultados obtenidos en las mismas en el Anexo III.

A la hora de presentar los resultados de la siguiente subsituación, en la misma tabla aparecen los modelos de las combinaciones donde la correlación entre el intercepto y la pendiente del nivel de aula es coincidente. De esta manera, los resultados de la combinación 1 y 2 aparecen en la Tabla VI.48, los resultados de la combinación 6 se presentan de manera individual (Tabla VI.49), los de la combinación 9, 10 y 11, en la Tabla VI.50 y los de la combinación 13, 14, 15 y 16 en la Tabla VI.51.

A continuación, se presentan los modelos de la **subsituación 19,** caracterizada porque el porcentaje que se fijó en la varianza del error intrasujeto es muy inferior a la suma de las varianzas del resto de coeficientes aleatorios. En los modelos de esta subsituación hay que recordar que el porcentaje que se fijó en la varianza del error intrasujeto, y los porcentajes totales de la varianza del intercepto y la pendiente, son los mismos que los fijados en el modelo 8, y en los distintos modelos de la subsituación 12, ya que, en los mismos se fijó la varianza de la pendiente del nivel de estudiante a un 47,5% de la varianza total del modelo, la varianza del intercepto de ese mismo nivel a otro 47,5% y la varianza del error intrasujeto, a un 5%.

En la tabla siguiente se muestra los resultados de los modelos de la subsituación 19, referidos a la **combinación 1 y 2**.

Tabla VI.48. *Matriz de correlaciones de las combinaciones 1 y 2 de la subsituación 19*

SUBSITUACIÓN 19

EFECTOS FIJOS		
Coef.		
$\boldsymbol{\beta_{000}}$	**4,6**	
$\boldsymbol{\beta_{100}}$	**0,4**	
EFECTOS ALEATORIOS		
NIVEL 3	$\sigma(u_{00j})$	$\sigma(u_{10j})$
$\boldsymbol{\sigma(u_{00j})}$	**23,75%**	
$\boldsymbol{\sigma(u_{10j})}$		**23,75%**
NIVEL 2	$\sigma(r_{0ij})$	$\sigma(r_{1ij})$
$\boldsymbol{\sigma(r_{0ij})}$	**23,75%**	
$\boldsymbol{\sigma(r_{1ij})}$		**23,75%**
NIVEL 1	$\sigma(e_{tij})$	
$\boldsymbol{\sigma(e_{tij})}$	**5%**	
Raulas y Restudiantes		
COMBINACIÓN 1		
$\boldsymbol{R_{aulas}}$ **: 0,2**		
$\boldsymbol{R_{estudiantes:}}$**:0,2**		
COMBINACIÓN 2		
$\boldsymbol{R_{aulas}}$**: 0,2**		
$\boldsymbol{R_{estudiantes}}$**:0,9**		

MATRIZ DE CORRELACIONES

COMBINACIÓN 1

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,7058**	**0,524**	**0,4245**	**0,364**	**0,3262**	**0,2997**	**0,2805**	**0,2663**
Y_{1ij}	0,7058	1	**0,935**	0,8986	0,8721	0,8532	0,8388	0,8281	0,8202
Y_{2ij}	0,524	0,935	1	0,9789	0,9697	0,9615	0,9546	0,9491	0,9449
Y_{3ij}	0,4245	0,8986	0,9789	1	0,9902	0,9872	0,9839	0,981	0,9786
Y_{4ij}	0,364	0,8721	0,9697	0,9902	1	0,9945	0,9933	0,9919	0,9905
Y_{5ij}	0,3262	0,8532	0,9615	0,9872	0,9945	1	0,9964	0,996	0,9953
Y_{6ij}	0,2997	0,8388	0,9546	0,9839	0,9933	0,9964	1	0,9975	0,9972
Y_{7ij}	0,2805	0,8281	0,9491	0,981	0,9919	0,996	0,9975	1	**0,9982**
Y_{8ij}	0,2663	0,8202	0,9449	0,9786	0,9905	0,9953	0,9972	0,9982	1

COMBINACIÓN 2

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,8185**	**0,7298**	**0,6767**	**0,6447**	**0,6212**	**0,6047**	**0,593**	**0,5842**
Y_{1ij}	0,8185	1	**0,9605**	0,9444	0,9321	0,9217	0,9142	0,9085	0,9042
Y_{2ij}	0,7298	0,9605	1	0,9858	0,9819	0,9777	0,9739	0,971	0,9687
Y_{3ij}	0,6767	0,9444	0,9858	1	0,9927	0,9913	0,9896	0,988	0,9867
Y_{4ij}	0,6447	0,9321	0,9819	0,9927	1	0,9955	0,9949	0,9942	0,9935
Y_{5ij}	0,6212	0,9217	0,9777	0,9913	0,9955	1	0,997	0,9968	0,9964
Y_{6ij}	0,6047	0,9142	0,9739	0,9896	0,9949	0,997	1	0,9978	0,9978
Y_{7ij}	0,593	0,9085	0,971	0,988	0,9942	0,9968	0,9978	1	0,9984
Y_{8ij}	0,5842	0,9042	0,9687	0,9867	0,9935	0,9964	0,9978	0,9984	1

Lo primero que se observa en los modelos de la tabla anterior, es un patrón decreciente entre la primera ocasión de medida y el resto de mediciones, tal y como ocurría en el modelo 8 y en los modelos de la subsituación 12.

Si se comparan los resultados del modelo de la combinación 1, con los del modelo 13, donde la correlación entre el estatus y la pendiente a nivel de aula se fijó a 0, y a nivel de estudiante a 0,2, se observa que los valores de correlación entre ambos modelos, son muy similares, casi idénticos, ya que, por ejemplo, en el modelo

12, la correlación entre la ocasión dos y tres era de 0,938, y en el modelo de la combinación 1 es de 0,935. Con nueve ocasiones de medida, dicha correlación es aun más similar entre ambos modelos, pues en el modelo 12, la correlacíon entre la ocasión ocho y nueve era de 0,9981 y en el modelo de la combinación 1 es de 0,9982.

Si se comparan los resultados del modelo de la combinación 2, con los del modelo 12.1, donde la correlación entre el intercepto y la pendiente del nivel de estudiante se fijó también a 0,9, la fiabilidad en las estimaciones de las puntuaciones de los estudiantes, es muy similar en ambos modelos. En la combinación 2, la correlación entre la ocasión dos y tres es de 0,960 y en el modelo 12.1 era de 0,978. En el modelo de la combinación 2, al igual que en el modelo 12.1, se sigue manteniendo el patrón decreciente entre la primera ocasión de medida y el resto de mediciones, a pesar de fijar una correlación de 0,9 entre el intercepto y la pendiente del nivel de estudiante.

En estos modelos de la combinación 1 y 2, al fijarse un porcentaje de varianza muy pequeño en el error intrasujeto, en comparación con el resto de coeficientes aleatorios, con tres ocasiones de medida se puede predecir el rendimiento de los estudiantes en los cursos posteriores de manera muy precisa.

En la siguiente tabla (Tabla VI. 49) se muestra el resultado de la **combinación 6,** donde se fijó, una correlación entre el intercepto y la pendiente del nivel de aula a 0,9, al igual que la del nivel de estudiante.

Tabla VI.49. *Matriz de correlaciones de la combinación 6 de la subsituación 19*

SUBSITUACIÓN 19

EFECTOS FIJOS

Coef.	
β_{000}	4,6
β_{100}	0,4

EFECTOS ALEATORIOS

NIVEL 3	$\sigma(u_{00j})$	$\sigma(u_{10j})$
$\sigma(u_{00j})$	23,75%	
$\sigma(u_{10j})$		23,75%

NIVEL 2	$\sigma(r_{0ij})$	$\sigma(r_{10j})$
$\sigma(r_{0ij})$	23,75%	
$\sigma(r_{10j})$		23,75%

NIVEL 1	$\sigma(e_{tij})$
$\sigma(e_{tij})$	5%

Raulas y Restudiantes

COMBINACIÓN 6

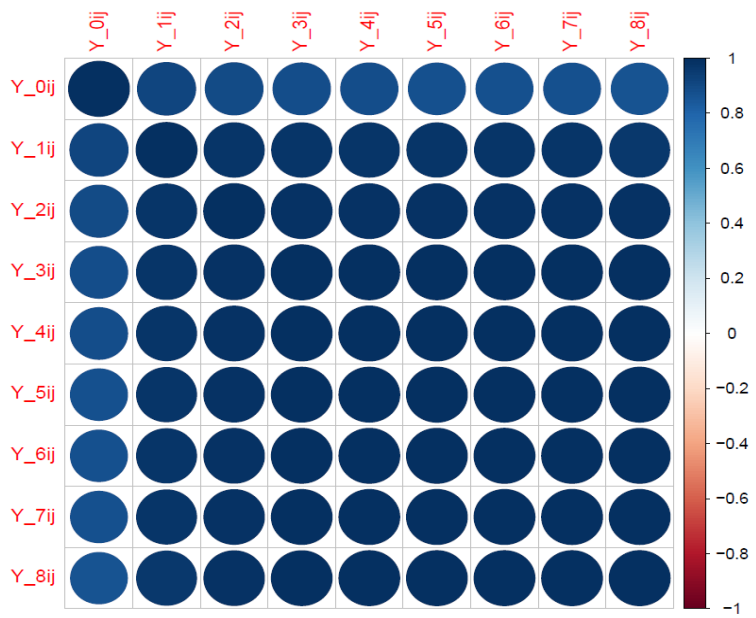
R_{aulas}: 0,9

$R_{estudiantes}$:0,9

MATRIZ DE CORRELACIONES

COMBINACIÓN 6

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,9136**	**0,9018**	**0,8926**	**0,8861**	**0,882**	**0,8784**	**0,8756**	**0,8733**
Y_{1ij}	0,9136	1	**0,9767**	0,9759	0,9745	0,9731	0,9719	0,9708	0,9698
Y_{2ij}	0,9018	0,9767	1	0,9892	0,9893	0,9891	0,9884	0,988	0,9876
Y_{3ij}	0,8926	0,9759	0,9892	1	0,9939	0,9941	0,9941	0,9939	0,9937
Y_{4ij}	0,8861	0,9745	0,9893	0,9939	1	0,996	0,9962	0,9962	0,9962
Y_{5ij}	0,882	0,9731	0,9891	0,9941	0,996	1	0,9972	0,9974	0,9974
Y_{6ij}	0,8784	0,9719	0,9884	0,9941	0,9962	0,9972	1	0,9979	0,9981
Y_{7ij}	0,8756	0,9708	0,988	0,9939	0,9962	0,9974	0,9979	1	0,9984
Y_{8ij}	0,8733	0,9698	0,9876	0,9937	0,9962	0,9974	0,9981	0,9984	1



Si se observa la combinación 6, se aprecia que entre la primera ocasión y el resto de mediciones se sigue manteniendo el patrón decreciente observado en las combinaciones anteriores, a pesar de haber fijado una correlación entre el intercepto y la pendiente del nivel de aula y del nivel de estudiante de 0,9.

Si se comparan los valores de correlación de este modelo con el modelo 12, donde también se fijó una correlación entre el estatus y la pendiente del nivel de estudiante de 0,9, ambos modelos muestran valores de correlación que son muy similares, así, por ejemplo, mientras que en el modelo 12.1, la correlación entre la ocasión dos y tres era de 0,978, en el modelo de la combinación 6 es de0,977.

En el gráfico de correlaciones, se puede observar que los valores de correlación más bajos se encuentran entre la primera ocasión de medida y el resto de mediciones, donde el color de los círculos es menos intenso. Pero a partir de esta primera ocasión de medida, las correlaciones son casi perfectas, por ello todos los círculos muestran un color azul intenso y un tamaño grande.

A continuación, se presentan los resultados de la **combinación 9, 10** y **11**. En estos modelos la correlación entre el intercepto y la pendiente del nivel de aula se fijó a -0,2.

Tabla VI.50. *Matriz de correlaciones de las combinaciones 9, 10 y 11, de la subsituación 19*

SUBSITUACIÓN 19

EFECTOS FIJOS	Coef.
β_{000}	4,6
β_{100}	0,4

EFECTOS ALEATORIOS		
NIVEL 3	$\sigma(u_{00j})$	$\sigma(u_{10j})$
$\sigma(u_{00j})$	23,75%	
$\sigma(u_{10j})$		23,75%
NIVEL 2	$\sigma(r_{0ij})$	$\sigma(r_{10j})$
$\sigma(r_{0ij})$	23,75%	
$\sigma(r_{10j})$		23,75%
NIVEL 1	$\sigma(e_{tij})$	
$\sigma(e_{tij})$	0,5%	

Raulas y Restudiantes

COMBINACIÓN 9

R_{aulas}: -0,2

$R_{estudiantes}$:0,2

COMBINACIÓN 10

R_{aulas}: -0,2

$R_{estudiantes}$:0,9

COMBINACIÓN 11

R_{aulas}: -0,2

$R_{estudiantes}$: -0,2

MATRIZ DE CORRELACIONES

COMBINACIÓN 9

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,6385**	**0,3842**	**0,2539**	**0,1848**	**0,1376**	**0,1071**	**0,0856**	**0,0685**
Y_{1ij}	0,6385	1	**0,9068**	0,8536	0,8187	0,7922	0,774	0,7606	0,75
Y_{2ij}	0,3842	0,9068	1	0,9728	0,961	0,9496	0,941	0,9342	0,9287
Y_{3ij}	0,2539	0,8536	0,9728	1	0,9887	0,9846	0,9807	0,9773	0,9741
Y_{4ij}	0,1848	0,8187	0,961	0,9887	1	0,9935	0,9919	0,9902	0,9884
Y_{5ij}	0,1376	0,7922	0,9496	0,9846	0,9935	1	0,996	0,9952	0,9944
Y_{6ij}	0,1071	0,774	0,941	0,9807	0,9919	0,996	1	0,9972	0,9969
Y_{7ij}	0,0856	0,7606	0,9342	0,9773	0,9902	0,9952	0,9972	1	0,9979
Y_{8ij}	0,0685	0,75	0,9287	0,9741	0,9884	0,9944	0,9969	0,9979	1

COMBINACIÓN 10

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,7478**	**0,6044**	**0,5251**	**0,4801**	**0,4484**	**0,4263**	**0,41**	**0,3968**
Y_{1ij}	0,7478	1	**0,9464**	0,9205	0,9017	0,887	0,8764	0,8681	0,8611
Y_{2ij}	0,6044	0,9464	1	0,9821	0,976	0,9699	0,9647	0,9606	0,9568
Y_{3ij}	0,5251	0,9205	0,9821	1	0,9916	0,9895	0,9871	0,9849	0,9827
Y_{4ij}	0,4801	0,9017	0,976	0,9916	1	0,9949	0,994	0,9929	0,9916
Y_{5ij}	0,4484	0,887	0,9699	0,9895	0,9949	1	0,9967	0,9963	0,9957
Y_{6ij}	0,4263	0,8764	0,9647	0,9871	0,994	0,9967	1	0,9976	0,9974
Y_{7ij}	0,41	0,8681	0,9606	0,9849	0,9929	0,9963	0,9976	1	0,9982
Y_{8ij}	0,3968	0,8611	0,9568	0,9827	0,9916	0,9957	0,9974	0,9982	1

COMBINACIÓN 11

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,5794**	**0,2652**	**0,1171**	**0,0379**	**-0,0109**	**-0,0424**	**-0,065**	**-0,0821**
Y_{1ij}	0,5794	1	0,8817	0,8169	0,774	0,7443	0,7239	0,7084	0,6969
Y_{2ij}	0,2652	0,8817	1	0,9684	0,9538	0,9416	0,9319	0,9243	0,9183
Y_{3ij}	0,1171	0,8169	0,9684	1	0,9871	0,9827	0,9784	0,9744	0,9712
Y_{4ij}	0,0379	0,774	0,9538	0,9871	1	0,993	0,9915	0,9895	0,9878
Y_{5ij}	-0,0109	0,7443	0,9416	0,9827	0,993	1	0,9957	0,9951	0,9941
Y_{6ij}	-0,0424	0,7239	0,9319	0,9784	0,9915	0,9957	1	0,997	0,9967
Y_{7ij}	-0,065	0,7084	0,9243	0,9744	0,9895	0,9951	0,997	1	0,9979
Y_{8ij}	-0,0821	0,6969	0,9183	0,9712	0,9878	0,9941	0,9967	0,9979	1

Tanto el modelo de la combinación 9, como los de la combinación 10 y 11, muestran un patrón decreciente entre la primera ocasión de medida y el resto de ocasiones.

Si se compara la matriz de datos de la combinación 9, con el modelo 12, los resultados son muy semejantes. Así, por ejemplo, en el modelo 12, la correlación entre la ocasión dos y tres era de 0,938 y en el modelo de la combinación 9 es de 0,907.

Si se compara el modelo 12.2, donde se fijó la correlación entre el estatus y la pendiente del nivel de estudiante a 0,9, con el modelo de la combinación 10, con el mismo número de ocasiones de medida, la precisión con la que se estima las puntuaciones de los estudiantes es muy parecida, ya que, en el modelo de la combinación 10, la correlación entre la ocasión dos y tres es de 0,946 y en el modelo 12.1 era de 0,978.

Comparando los resultados de la combinación 11, con los del modelo 12.2, donde también se fijó una correlación entre el intercepto y la pendiente del nivel de estudiante de -0,2, en ambos modelos se obtienen resultados muy similares, tanto que en los dos modelos la correlación entre la ocasión ocho y nueve es de 0,998.

Si se comparan los resultados de la combinación 9, 10 y 11, aunque los tres modelos muestran valores de correlación muy similares, son inferiores en el modelo 11, sobre todo con pocas mediciones. Esto se debe a que, en dicho modelo, se fijó una correlación negativa no solo entre el intercepto y la pendiente del nivel de aula, sino también, entre el estatus y la pendiente del nivel de estudiante.

Finalmente, dentro de esta subsituación 19, se presentan los modelos de las **combinaciones 13, 14, 15** y **16**, donde la correlación entre el estatus y la pendiente del nivel de aula siempre se fijó a -0,9.

Tabla VI.51. *Matriz de correlaciones de las combinaciones 13, 14, 15 y 16, de la subsituación 19*

SUBSITUACIÓN 19

EFECTOS FIJOS	Coef.
β_{000}	4,6
β_{100}	0,4
EFECTOS ALEATORIOS	
NIVEL 3	$\sigma(u_{00j})$ $\sigma(u_{10j})$
$\sigma(u_{00j})$	1,25%
$\sigma(u_{10j})$	1,25%
NIVEL 2	$\sigma(r_{0ij})$ $\sigma(r_{1ij})$
$\sigma(r_{0ij})$	1,25%
$\sigma(r_{1ij})$	1,25%
NIVEL 1	$\sigma(e_{tij})$
$\sigma(e_{tij})$	95%

Raulas y Restudiantes

COMBINACIÓN 13

R_{aulas}: -0,9

$R_{estudiantes}$:0,2

COMBINACIÓN 14

R_{aulas}: -0,9

$R_{estudiantes}$:0,9

COMBINACIÓN 15

R_{aulas}: -0,9

$R_{estudiantes}$: -0,2

COMBINACIÓN 16

R_{aulas}: -0,9

$R_{estudiantes}$: -0,9

MATRIZ DE CORRELACIONES

COMBINACIÓN 13

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,5189**	**0,1379**	**-0,033**	**-0,12**	**-0,1687**	**-0,2013**	**-0,2249**	**-0,2416**
Y_{1ij}	0,5189	1	**0,8534**	0,7701	0,7169	0,6839	0,6605	0,6428	0,6303
Y_{2ij}	0,1379	0,8534	1	0,9631	0,9458	0,9317	0,9211	0,9124	0,906
Y_{3ij}	-0,033	0,7701	0,9631	1	0,986	0,9812	0,9764	0,9721	0,9686
Y_{4ij}	-0,12	0,7169	0,9458	0,986	1	0,9929	0,9911	0,9891	0,9872
Y_{5ij}	-0,1687	0,6839	0,9317	0,9812	0,9929	1	0,9957	0,9948	0,9938
Y_{6ij}	-0,2013	0,6605	0,9211	0,9764	0,9911	0,9957	1	0,997	0,9967
Y_{7ij}	-0,2249	0,6428	0,9124	0,9721	0,9891	0,9948	0,997	1	0,9979
Y_{8ij}	-0,2416	0,6303	0,906	0,9686	0,9872	0,9938	0,9967	0,9979	1

COMBINACIÓN 14

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	1	**0,6753**	**0,4483**	**0,3275**	**0,2581**	**0,2121**	**0,1816**	**0,1598**	**0,1419**
Y_{1ij}	0,6753	1	0,9178	0,8701	0,836	0,8106	0,7932	0,78	0,7688
Y_{2ij}	0,4483	0,9178	1	0,9747	0,9631	0,9522	0,9436	0,9367	0,9307
Y_{3ij}	0,3275	0,8701	0,9747	1	0,9886	0,9846	0,9805	0,9768	0,9735
Y_{4ij}	0,2581	0,836	0,9631	0,9886	1	0,9936	0,992	0,9902	0,9882
Y_{5ij}	0,2121	0,8106	0,9522	0,9846	0,9936	1	0,9959	0,9953	0,9943
Y_{6ij}	0,1816	0,7932	0,9436	0,9805	0,992	0,9959	1	0,9973	0,9969
Y_{7ij}	0,1598	0,78	0,9367	0,9768	0,9902	0,9953	0,9973	1	0,9979
Y_{8ij}	0,1419	0,7688	0,9307	0,9735	0,9882	0,9943	0,9969	0,9979	1

COMBINACIÓN 15

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	**1**	**0,4296**	**-0,0537**	**-0,2308**	**-0,318**	**-0,3629**	**-0,3937**	**-0,4137**	**-0,4284**
Y_{1ij}	0,4296	1	0,7877	0,6886	0,6274	0,5918	0,5664	0,5495	0,5367
Y_{2ij}	-0,0537	0,7877	1	0,9558	0,9379	0,9241	0,9133	0,9053	0,899
Y_{3ij}	-0,2308	0,6886	0,9558	1	0,9834	0,9789	0,9742	0,9703	0,967
Y_{4ij}	-0,318	0,6274	0,9379	0,9834	1	0,9918	0,9903	0,9887	0,9869
Y_{5ij}	-0,3629	0,5918	0,9241	0,9789	0,9918	1	0,9949	0,9945	0,9936
Y_{6ij}	-0,3937	0,5664	0,9133	0,9742	0,9903	0,9949	1	0,9967	0,9964
Y_{7ij}	-0,4137	0,5495	0,9053	0,9703	0,9887	0,9945	0,9967	1	0,9976
Y_{8ij}	-0,4284	0,5367	0,899	0,967	0,9869	0,9936	0,9964	0,9976	1

COMBINACIÓN 16

	Y_{0ij}	Y_{1ij}	Y_{2ij}	Y_{3ij}	Y_{4ij}	Y_{5ij}	Y_{6ij}	Y_{7ij}	Y_{8ij}
Y_{0ij}	**1**	**0,4296**	**-0,0537**	**-0,2308**	**-0,318**	**-0,3629**	**-0,3937**	**-0,4137**	**-0,4284**
Y_{1ij}	0,4296	1	0,7877	0,6886	0,6274	0,5918	0,5664	0,5495	0,5367
Y_{2ij}	-0,0537	0,7877	1	0,9558	0,9379	0,9241	0,9133	0,9053	0,899
Y_{3ij}	-0,2308	0,6886	0,9558	1	0,9834	0,9789	0,9742	0,9703	0,967
Y_{4ij}	-0,318	0,6274	0,9379	0,9834	1	0,9918	0,9903	0,9887	0,9869
Y_{5ij}	-0,3629	0,5918	0,9241	0,9789	0,9918	1	0,9949	0,9945	0,9936
Y_{6ij}	-0,3937	0,5664	0,9133	0,9742	0,9903	0,9949	1	0,9967	0,9964
Y_{7ij}	-0,4137	0,5495	0,9053	0,9703	0,9887	0,9945	0,9967	1	0,9976
Y_{8ij}	-0,4284	0,5367	0,899	0,967	0,9869	0,9936	0,9964	0,9976	1

Lo primero que se observa en estos modelos de la tabla anterior VI. 51, es que, la correlación entre la primera ocasión de medida y el resto de ocasiones se sigue manteniendo decreciente.

Si se comparan los resultados de la combinación 13 con los del modelo 12, se obtiene valores de correlación muy semejantes en ambos modelos, por ejemplo, en el modelo 12, la correlación entre la ocasión tres y cuatro era de 0,979 y, en el modelo de la combinación 13 es de 0,961. También el modelo 12.1 muestra valores de correlación muy semejantes al del modelo de la combinación 14, y lo mismo pasa si se compara el modelo de la combinación 15, con el modelo 12.2, y el modelo de la combinación 16, con el modelo 12.3, ya que, por ejemplo, en dicho modelo la correlación entre la ocasión dos y tres era de 0,458 y en el modelo de la combinación 16 es de 0,462.

Mediante los resultados obtenidos en los modelos de la subsituación 19, se puede concluir algo muy similar a lo comentado para los modelos de las distintas combinaciones de las subsituaciones anteriores, y, es que, cuando se comparan dos modelos en los cuales se fija el mismo valor de correlación entre el intercepto y la pendiente de uno de los niveles, pero en el otro nivel, un modelo cuenta con una correlación determinada y en el otro modelo no se fija correlación, las puntuaciones de los estudiantes se estiman con una precisión similar en ambos modelos, aunque en uno de ellos no se establezca relación entre el rendimiento de partida y el que se desarrolla a lo largo del tiempo.

Dicha similitud puede ser debida, a que tal y como se ha visto en los modelos de *la situación b*, el valor de correlación fijado entre el intercepto y la pendiente no tiene un efecto tan marcado en la fiabilidad con la que se estima las puntuaciones de los estudiantes, como lo puede tener el porcentaje de varianza fijado en la pendiente de crecimiento y/o en el error intrasujeto. Esto unido a que, en los modelos comparados entre las distintas subsituaciones se fijó en uno de los niveles, el mismo valor de correlación entre el intercepto y la pendiente, y los porcentajes de varianza totales también eran los mismos, hace que las matrices de correlaciones de los modelos que comparten estas características sean muy similares.

Finalmente, como colofón a este primer problema de investigación, se presentan los resultados obtenidos acorde a la pregunta, **¿qué aulas potencian más el rendimiento de los estudiantes agrupados en los distintos quintiles?**

VI.1.2.4. Estudio del rendimiento de los estudiantes agrupados en los distintos quintiles

Para estudiar el rendimiento de los estudiantes en los diferentes quintiles, recuérdese, que se utilizaron los datos generados en cada uno de los cuatro modelos de la **subsituación 14** (modelo 14, modelo 14.1, modelo 14.2 y modelo 14.3), ya que, no había influencia de las características de los estudiantes en la pendiente de crecimiento de las aulas. En dichos modelos, se fijó el porcentaje de varianza del intercepto del nivel de aula a un 55%, la varianza de la pendiente de dicho nivel a un 25%, la varianza del estatus del nivel de estudiante a un 15%, y la varianza del error intrasujeto a un 5%, así como distintos valores de correlación entre el intercepto y la pendiente del nivel de aula. En cada uno de los cuatro modelos, se estudian los resultados referidos a los tres tipos de aula (con pendiente baja, media y alta), seleccionadas en cada uno de los cinco quintiles, que a su vez se encuentran compuestas por estudiantes pertenecientes a cada uno de los quintiles.

En cada una de las tablas de resultados referidas a cada uno de los cinco quintiles de los cuatro modelos de la subsituación 14, (desde la Tabla VI.52 hasta la VI.71), se especifica el número de aula (aula) y dentro de la misma, los quintiles en los que se agrupan los estudiantes ($\mathrm{Z(r_{0ij})\_}q$), el número de estudiantes por quintil (N), la puntuación promedia de los estudiantes en la primera ocasión de medida (Y_{0ij}) y en la última (Y_{8ij}), la diferencia promedia entre la puntuación de los estudiantes obtenida en la última ocasión y la primera ($Y_{8ij} - Y_{0ij}$), señalando de color rojo los estudiantes que menos crecen y de azul los que más, los coeficientes aleatorios del nivel de aula y estudiante tipificados ($Z\mathrm{u}_{00j}$, $Z\mathrm{u}_{10j}$, Zr_{0ij}) y sin tipificar(u_{00j}, u_{10j}, r_{0ij}), aunque se interpretan los tipificados para facilitar la comprensión, y finalmente los residuos del error intrasujeto en la primera y en la última ocasión de medida ($\mathrm{e}_{0ij}, \mathrm{e}_{8ij}$). Además, dichas tablas se acompañan de un gráfico en el que se representa, de manera más visual, la diferencia promedia en las puntuaciones de los estudiantes, entre la primera ocasión de medida y la última,

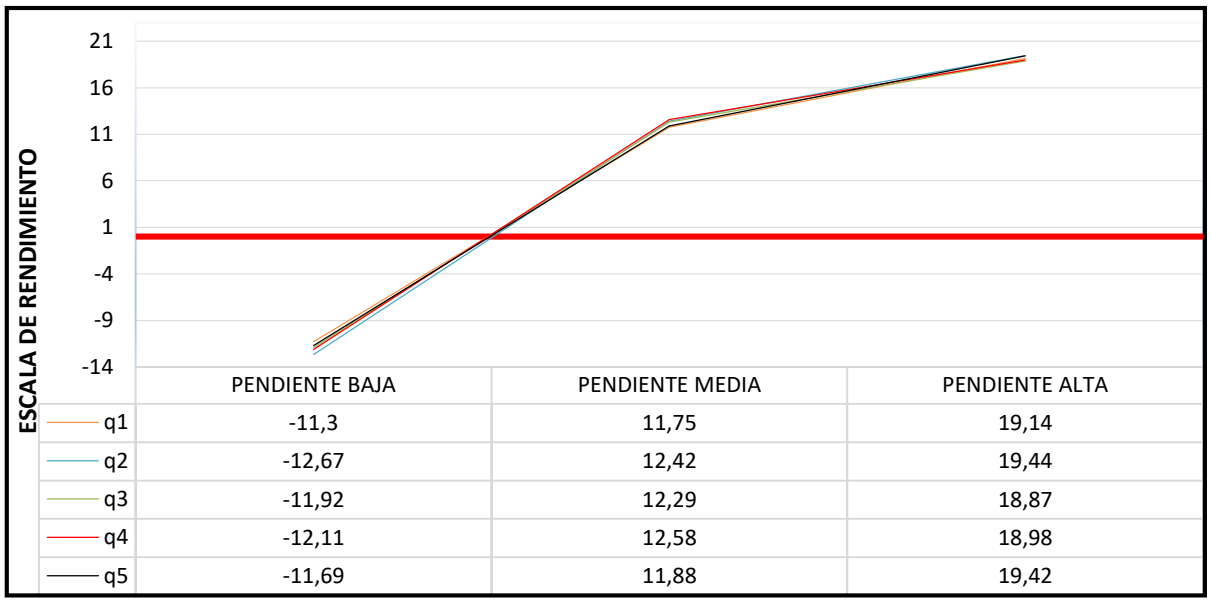
acorde al quintil al que pertenezcan y al tipo de aula al que asistan (pendiente baja, media o alta).

A continuación, se presentan los resultados de los cinco quintiles obtenidos a partir del **modelo 14**, en el cual se fijó una correlación entre el intercepto y la pendiente del nivel de aula, baja y positiva (0,2).

Concretamente, en la tabla siguiente VI.52 se muestra los resultados de los estudiantes que asisten a aulas que por su valor promedio en el punto de partida se agrupan en el primer quintil, es decir, son aulas que parten de un rendimiento previo muy bajo, aunque están formadas por estudiantes agrupados en los cinco quintiles.

Tabla VI.52. *Resultados del modelo 14 quintil 1*

MODELO 14 QUINTIL 1													
AULA	$Z(r_{0ij})\_q$	**N**	Y_{0ij}	Y_{8ij}	$Y_{8ij} - Y_{0ij}$	$Z(u_{00j})$	$Z(u_{10j})$	$Z(r_{0ij})$	u_{00j}	u_{10j}	r_{0ij}	e_{0ij}	e_{8ij}
266	**1**	5	-0,64	-11,94	**-11,3**	-1,39	**-1,18 (Baja)**	-1,38	-3,03	-1,92	-1,67	-0,55	0,33
	2	4	1,2	-11,47	**-12,67**			-0,52			-0,64	0,27	-0,23
	3	6	1,56	-10,36	-11,92			-0,05			-0,07	0,06	0,31
	4	3	2,36	-9,75	-12,11			0,59			0,7	0,08	0,15
	5	7	3,2	-8,49	-11,69			1,27			1,52	0,11	0,59
286	**1**	9	-0,03	11,72	**11,75**	-1,42	**0,62 (Media)**	-1,22	-3,11	1,11	-1,49	-0,04	-0,4
	2	2	0,22	12,64	12,42			-0,72			-0,88	-0,39	-0,09
	3	5	1,13	13,42	12,29			0			-0,01	-0,36	-0,19
	4	4	1,82	14,4	**12,58**			0,55			0,65	-0,32	0,14
	5	5	3,34	15,22	11,88			1,34			1,61	0,24	0
173	1	4	0,33	19,47	19,14	-1,37	**1,17 (Alta)**	-1,19	-2,98	2,05	-1,44	0,16	-0,3
	2	1	1,18	20,62	**19,44**			-0,48			-0,59	0,16	-0,01
	3	4	1,74	20,61	**18,87**			0,11			0,12	0	-0,73
	4	1	2,26	21,24	18,98			0,31			0,36	0,28	-0,33
	5	7	3,36	22,78	19,42			1,46			1,76	-0,02	-0,19



Como se puede observar en las aulas del quintil uno, los estudiantes que pertenecen a un aula que presenta una pendiente baja y negativa, por debajo de la media (-1,18), como es el aula 266, independientemente del quintil al que pertenezcan, no incrementan su rendimiento desde la primera ocasión de medida a la última, por ello, en el gráfico los valores de estos estudiantes se encuentran por debajo de 0.

En promedio, todos los estudiantes que asisten al aula 266, ven como su rendimiento disminuye con el paso del tiempo, entre 11 y 13 puntos. Los estudiantes que se encuentran en el quintil dos, son los que muestran un mayor descenso de su tasa de crecimiento, ya que, parten de un rendimiento promedio de 1,2 puntos, y en la ocasión nueve, su rendimiento es de -11,47, por lo tanto, su rendimiento ha disminuido 12,67 puntos. Sin embargo, los estudiantes situados en el primer quintil son los que muestran un menor descenso en su pendiente de crecimiento, esto puede ser debido, a que, aunque parten de un rendimiento promedio bajo y negativo, asisten a aulas donde la mayoría de los estudiantes cuentan con características semejantes.

Referido a los estudiantes situados en el quintil cinco, aunque parten de un rendimiento promedio más alto al del resto de estudiantes, con el paso del tiempo tienen una ganancia negativa que es inferior a la de los estudiantes que se sitúan en el resto de quintiles, a excepción de los del quintil uno. Así, en esta aula, tanto los estudiantes que parten de un nivel promedio que es muy similar al rendimiento promedio de partida del aula al que asisten (quintil uno), como, los estudiantes que parten de un rendimiento promedio muy superior al logro promedio del aula al que pertenecen (quintil cinco), son los estudiantes que menos ven afectado su tasa de rendimiento cuando asisten a un aula situada en el quintil uno, con una pendiente de crecimiento baja y negativa.

En cuanto a los estudiantes que pertenecen a un aula, como es la 286, donde la pendiente de crecimiento es media (0,62), se observa que los estudiantes agrupados en cada uno de los cinco quintiles ven incrementado su rendimiento con el paso del tiempo, tal y como, también se puede apreciar en el gráfico, ya que, las líneas de los estudiantes categorizados en los cinco quintiles se encuentran por encima de 0. Todos los estudiantes en promedio incrementan su rendimiento desde la primera ocasión de medida a la última entre unos 11 y 12,5 puntos. En esta aula los estudiantes que pertenecen al quintil uno, son los que menos incrementan su

rendimiento con el paso del tiempo (11,75 puntos), seguidos de los estudiantes del quintil cinco, que su rendimiento se incrementa a lo largo del tiempo 11,88 puntos. Lo contrario a lo que ocurría en el aula 266, donde eran estos estudiantes situados en dichos quintiles, los menos perjudicados por asistir a un aula con una pendiente baja y negativa.

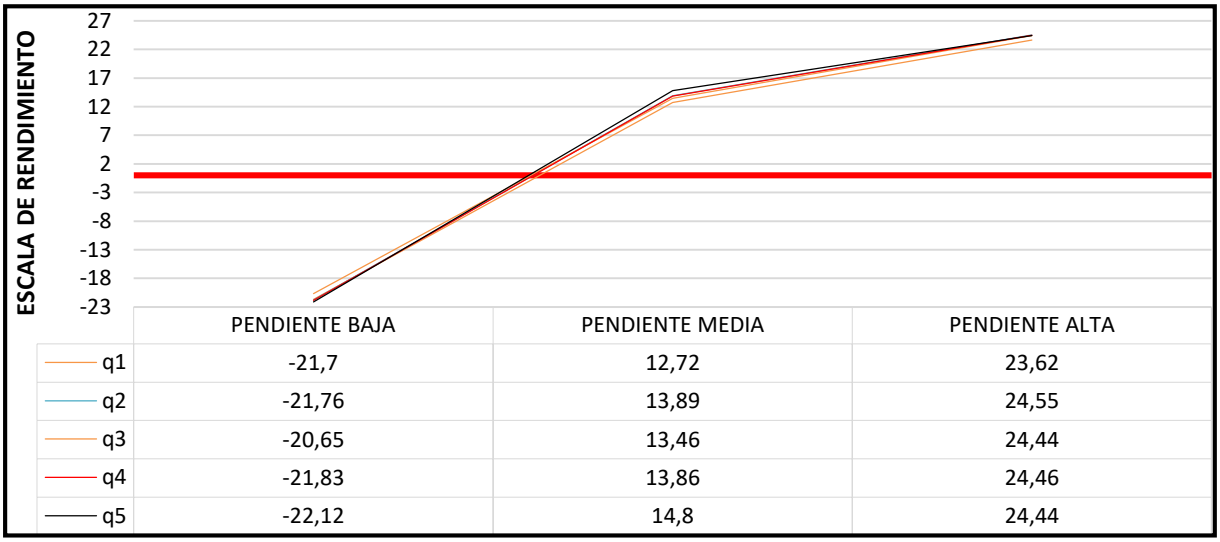
Finalmente, en cuanto a los estudiantes que asisten al aula 173, caracterizada por tener una pendiente de crecimiento promedia alta (1,17), se observa que el rendimiento de los estudiantes situados en los diferentes quintile se incrementa entre 18 y 19,5 puntos entre la primera ocasión de medida y la última. Dicha tasa de crecimiento es superior en los estudiantes que se encuentran situados en el segundo quintil y en el quintil cinco, debido a que incrementan su rendimiento más de 19 puntos. En cuanto a los estudiantes que menos crecen desde la ocasión una a la última, son los del quintil tres, puesto que, su rendimiento se incrementa en 18,87 en la escala de rendimiento.

A la vista de estos resultados se puede concluir que, los estudiantes situados en el primer quintil, excepto cuando la pendiente del aula es media, son los que presentan resultados más positivos al asistir a un aula que está formado en su mayoría por estudiantes que cuentan con un rendimiento de partida bajo (aulas también situadas en el primer quintil). En cuanto a los estudiantes pertenecientes al quintil cinco, a pesar de asistir a aulas en las cuales, la mayoría de los estudiantes parten de un rendimiento promedio bajo, su tasa de crecimiento no se ve afectada de manera muy negativa, en comparación con la tasa de crecimiento de los estudiantes situados en el resto de quintiles. Además, la tasa de crecimiento de los estudiantes pertenecientes a una misma aula, pero agrupados en distinto quintil, no muestra diferencias muy marcadas entre sí, pues, como se observa en el gráfico las líneas se encuentran sobrepuestas.

A continuación, se presentan los resultados de las aulas seleccionadas dentro del quintil dos del modelo 14.

Tabla VI.53. *Resultados del modelo 14 quintil 2*

MODELO 14 QUINTIL 2													
AULA	$Z(r_{0ij})\_q$	N	Y_{0ij}	Y_{8ij}	$Y_{8ij} - Y_{0ij}$	$Z(u_{00j})$	$Z(u_{10j})$	$Z(r_{0ij})$	u_{00j}	u_{10j}	r_{0ij}	e_{0ij}	e_{8ij}
215	1	6	1,8	-19,9	-21,7	-0,61	**-1,88 (Baja)**	-1,52	-1,28	-3,1	-1,84	0,32	0,24
	2	6	3,07	-18,69	-21,76			-0,52			-0,64	0,38	0,24
	3	5	2,86	-17,79	-20,65			0,04			0,03	-0,5	0,47
	4	3	4,25	-17,58	-21,83			0,49			0,59	0,34	0,13
	5	2	5,15	-16,97	-22,12			1,18			1,41	0,42	-0,09
248	1	4	2,13	14,85	12,72	-0,62	**0,74 (Media)**	-1,5	-1,3	1,32	-1,82	0,65	-0,36
	2	5	2,67	16,56	13,89			-0,64			-0,79	0,16	0,32
	3	10	3,61	17,07	13,46			0			-0,01	0,32	0,05
	4	6	4,35	18,21	13,86			0,57			0,68	0,37	0,49
	5	3	4,34	19,14	14,8			1,29			1,55	-0,51	0,55
185	1	1	1,58	25,2	23,62	-0,62	**1,5 (Alta)**	-1,11	-1,29	2,61	-1,35	-0,38	-0,86
	2	7	2,27	26,82	24,55			-0,53			-0,64	-0,4	0,06
	3	4	3,18	27,62	24,44			0,06			0,07	-0,2	0,15
	4	8	4,01	28,47	24,46			0,57			0,67	0,02	0,39
	5	5	4,98	29,42	24,44			1,4			1,68	-0,01	0,34



Como se observa en la tabla anterior referida a las aulas del quintil 2, los estudiantes que asisten a aulas con una pendiente de crecimiento negativa (-1,88), con el paso del tiempo su tasa de crecimiento tiene una ganancia negativa comprendida entre unos 20 y 23 puntos, en la escala de rendimiento. Siendo los estudiantes situados en el quintil tres, los que menos afectados se encuentran por dicha pérdida, ya que, su logro se ve disminuido unos 20,65 puntos. Por el contrario, son los estudiantes del quintil cinco los que más ven disminuido su rendimiento desde la primera ocasión a la última (22,12 puntos), esto puede ser debido a que, asisten a un aula que en promedio presenta un rendimiento muy bajo. Aunque, tal y como se observa en el gráfico, las diferencias entre los estudiantes situados en los

diversos quintiles no son muy marcadas, sólo entre los estudiantes pertenecientes al quintil tres y cinco se aprecia una mayor desigualdad.

En cuanto a los estudiantes que asisten al aula 248, cuya pendiente de crecimiento, en puntuaciones tipificadas, es de 0,74 (pendiente media), se observa que en promedio los estudiantes de los distintos quintiles crecen entre unos 12-15 puntos a lo largo del tiempo. Siendo los estudiantes situados en el quintil cinco los que más crecen (14,8) y los del quintil uno los que menos (12,72). Por ello, en el gráfico las líneas que representan a ambos quintiles, dentro del aula con pendiente media, son las que aparecen más distanciadas, mostrando un rendimiento superior los estudiantes situados en el quintil cinco.

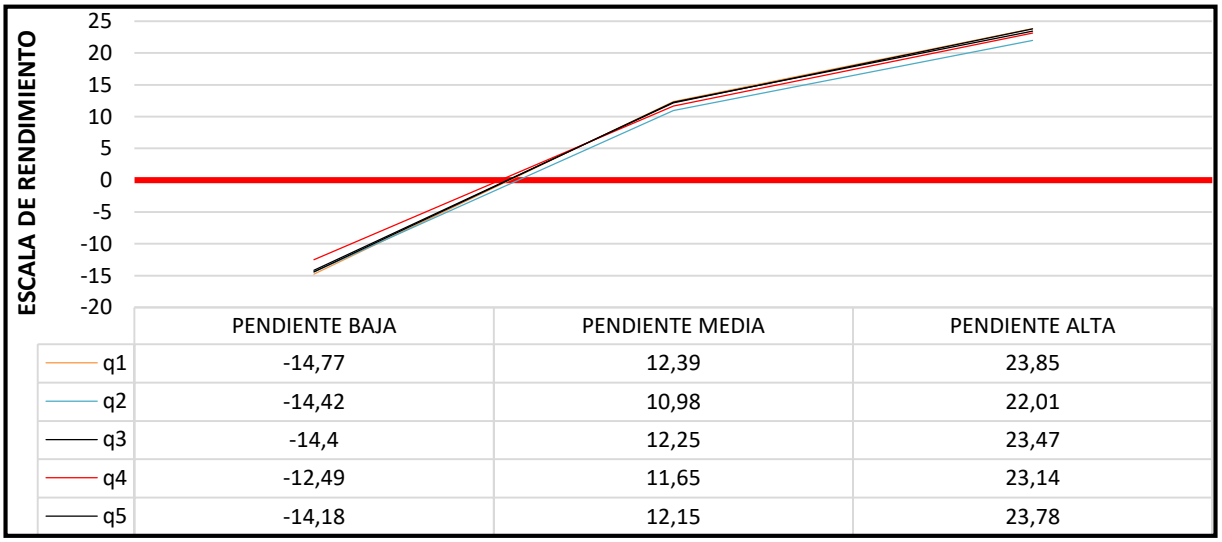
Finalmente, en cuanto a los estudiantes que acuden al aula 185, donde su pendiente de crecimiento promedia es alta (1,5) se observa que los estudiantes que menos crecen son los del quintil uno. Los estudiantes situados en el resto de quintiles crecen de manera muy homogénea (unos 24,5 puntos), por ello, en el gráfico las líneas aparecen muy juntas.

A la vista de estos resultados, en general, los estudiantes pertenecientes al quintil uno, son los que menos incrementan su tasa de crecimiento cuando asisten a un aula situada en el quintil dos, excepto cuando la pendiente de dicha aula es baja. Lo contrario ocurre con los estudiantes situados en el quintil cinco, ya que, son los estudiantes que presentan una mejor tasa de crecimiento, menos cuando asisten al aula 215, donde la pendiente es baja y negativa.

A continuación, se presentan los resultados obtenidos en los estudiantes que asisten a un aula, que por su rendimiento promedio de partida se encuentran situadas en el tercer quintil.

Tabla VI.54. *Resultados del modelo 14 quintil 3*

MODELO 14 QUINTIL 3													
AULA	$Z(r_{0ij})$ _q	N	Y_{0ij}	Y_{8ij}	$Y_{8ij} - Y_{0ij}$	$Z(u_{00j})$	$Z(u_{10j})$	$Z(r_{0ij})$	u_{00j}	u_{10j}	r_{0ij}	e_{0ij}	e_{8ij}
90	1	5	3,92	-10,85	-14,77	0,11	**-1,32 (Baja)**	-1,27	0,34	-2,17	-1,55	0,53	-0,11
	2	7	4,33	-10,09	-14,42			-0,57			-0,7	0,09	-0,2
	3	7	5,06	-9,34	-14,4			0,05			0,05	0,07	-0,2
	4	2	4,75	-7,74	-12,49			0,46			0,55	-0,74	0,9
	5	3	6,72	-7,46	-14,18			1,35			1,62	0,15	0,11
384	1	6	3,43	15,82	12,39	0,11	**0,62 (Media)**	-1,31	0,35	1,12	-1,59	0,07	0,29
	2	3	4,88	15,86	10,98			-0,51			-0,62	0,55	-0,65
	3	5	5,16	17,41	12,25			0,01			0	0,21	0,28
	4	8	5,87	17,52	11,65			0,55			0,66	0,26	-0,26
	5	4	6,02	18,17	12,15			1,14			1,37	-0,3	-0,32
261	1	3	2,03	25,88	23,85	0,12	**1,43 (Alta)**	-2,02	0,37	2,5	-2,45	-0,49	0,19
	2	6	4,77	26,78	22,01			-0,56			-0,69	0,49	-0,67
	3	8	5,04	28,51	23,47			0,02			0,02	0,06	0,35
	4	2	5,86	29	23,14			0,58			0,69	0,21	0,17
	5	8	6,32	30,1	23,78			1,25			1,5	-0,15	0,46



Cuando los estudiantes pertenecen a un aula, como es el aula 90, en la cual la pendiente es baja (-1,32), todos, independientemente del quintil en el que se sitúen, tienen una ganancia negativa, por ello en el gráfico aparecen por debajo de la media. Son los estudiantes situados en el primer quintil los que más ven disminuido su tasa de crecimiento (14,77 puntos), muy seguidos de los del quintil dos (14,42), quintil tres (14,4), y quintil cinco (14,18). Sin embargo, los que menos disminuyen su tasa de crecimiento son los estudiantes situados en el quintil cuatro (12,49).

Los estudiantes que asisten al aula 384, obtienen una ganancia con el paso del tiempo entre 10 y 12,5 puntos, siendo los estudiantes que se encuentran en el quintil uno, los que en promedio más incrementan su logro. Lo contrario a lo que ocurría en

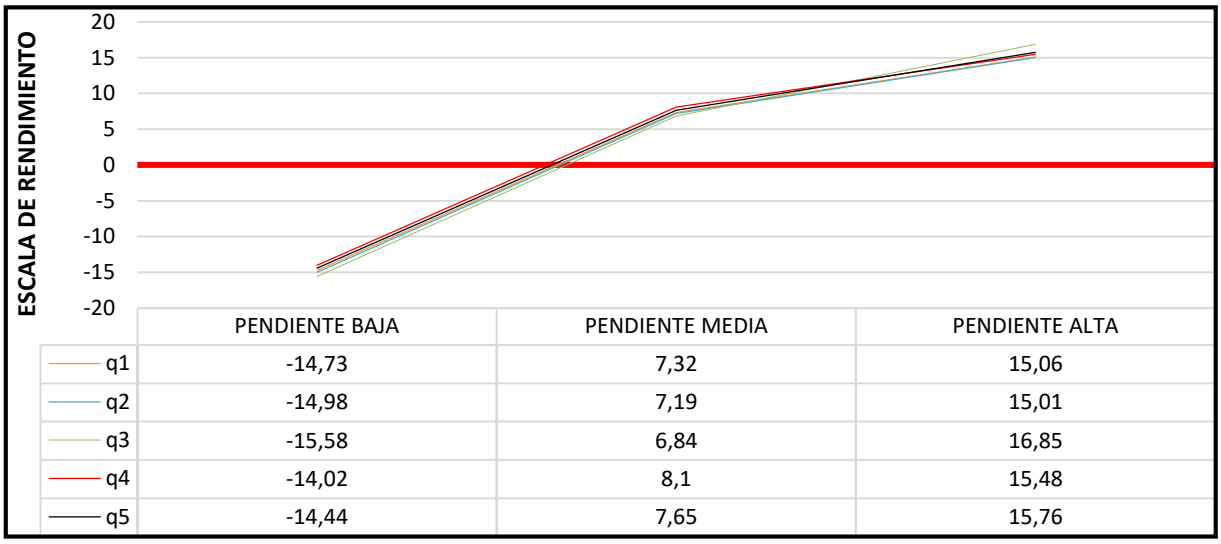
el aula 90, por ello, en el gráfico se observa como los estudiantes de este quintil, pasan de situarse en la parte más baja cuando la pendiente es baja y en la parte más alta cuando la pendiente es media, manteniéndose dicho posicionamiento incluso cuando la pendiente es alta, ya que, en este contexto también son los estudiantes situados en el quintil uno los que muestran una mayor tasa de crecimiento (23,85 puntos).Por el contrario, son los estudiantes situados en el quintil dos, los que muestran una menor ganancia en la tasa de rendimiento entre la primera ocasión y la última, tanto cuando la pendiente del aula es media, como cuando es alta.

A la vista de estos resultados parece que, excepto cuando la pendiente del aula es baja y negativa, los estudiantes pertenecientes al quintil uno son los que más incrementan su tasa de crecimiento, en comparación con los estudiantes situados en el resto de quintiles. Sin embargo, son los estudiantes situados en el segundo quintil los que, en general, menos incrementan su tasa de crecimiento cuando asisten a un aula situada en el tercer quintil, ya que, son los que menos ganancia obtienen.

En la siguiente tabla, se muestran los resultados de las aulas que han sido seleccionadas dentro del cuarto quintil.

Tabla VI.55. *Resultados del modelo 14 quintil 4*

MODELO 14 QUINTIL 4													
AULA	$Z(r_{0ij})$ _q	**N**	Y_{0ij}	Y_{8ij}	$Y_{8ij} - Y_{0ij}$	$Z(u_{00j})$	$Z(u_{10j})$	$Z(r_{0ij})$	u_{00j}	u_{10j}	r_{0ij}	e_{0ij}	e_{8ij}
14	1	8	4,25	-10,48	-14,73	0,58	**-1,42 (Baja)**	-1,26	1,41	-2,33	-1,54	-0,23	0,47
	2	2	5,23	-9,75	-14,98			-0,62			-0,76	-0,02	0,42
	3	9	6,16	-9,42	**-15,58**			0,05			0,05	0,1	-0,05
	4	3	5,79	-8,23	**-14,02**			0,59			0,7	-0,91	0,49
	5	3	6,68	-7,76	-14,44			1,03			1,24	-0,57	0,41
113	1	8	4,7	12,02	7,32	0,58	**0,25 (Media)**	-1,15	1,4	0,5	-1,4	0,1	0,21
	2	6	5,35	12,54	7,19			-0,51			-0,62	-0,03	-0,04
	3	4	6,13	12,97	**6,84**			0,09			0,11	0,03	-0,34
	4	2	6,17	14,27	**8,1**			0,56			0,66	-0,5	0,4
	5	3	7,05	14,7	7,65			1,32			1,58	-0,54	-0,09
178	1	8	4,71	19,77	15,06	0,58	**0,86 (Alta)**	-1,29	1,42	1,52	-1,57	0,26	-0,01
	2	1	5,49	20,5	**15,01**			-0,64			-0,78	0,25	-0,07
	3	1	3,47	20,32	**16,85**			-0,2			-0,26	-2,29	-0,78
	4	6	6,84	22,32	15,48			0,51			0,61	0,21	0,35
	5	9	7,61	23,37	15,76			1,44			1,73	-0,13	0,29



Como se puede observar en la tabla VI.55, los estudiantes que asisten a un aula con una pendiente baja y negativa como es el aula 14, ven disminuido su tasa de crecimiento entre 14 y 15,6 puntos; siendo los estudiantes situados en el quintil tres, los que muestran una mayor pérdida y los del quintil cuatro, los que menos.

Este patrón también se mantiene cuando los estudiantes asisten a un aula donde la pendiente es media (aula 113). En esta aula, aunque los estudiantes de todos los quintiles incrementan su rendimiento entre 7 y 8 puntos, nuevamente son los estudiantes del quintil tres los que menos incrementan su tasa de crecimiento, mientras que, los estudiantes del quintil cuatro, son los que más crecen. Por ello, en el gráfico se observa que cuando la pendiente es baja y media, la línea

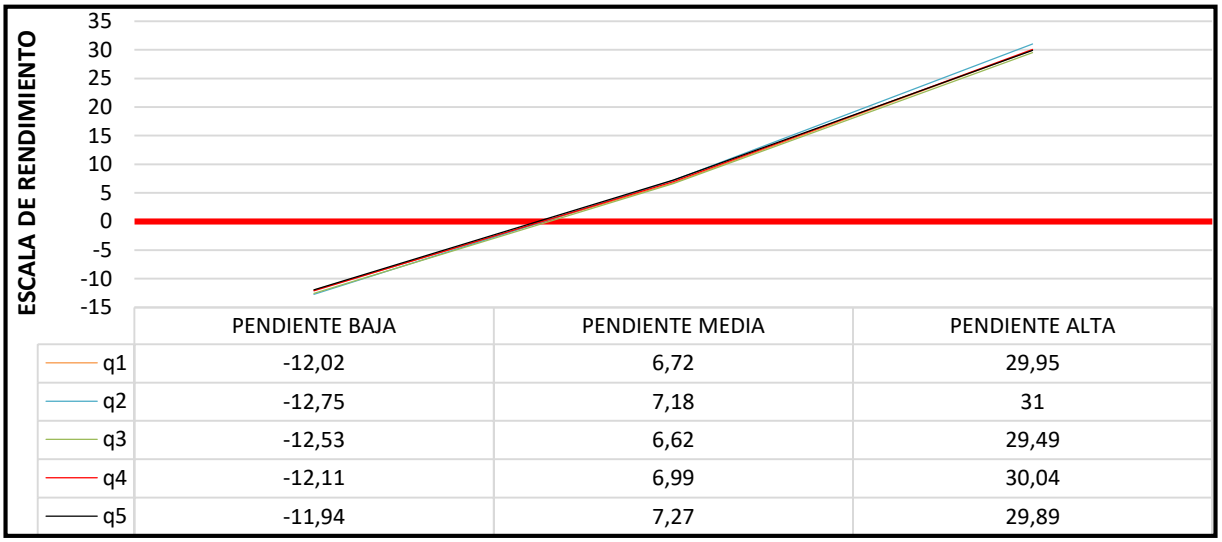
correspondiente al quintil cuatro se encuentra en la parte superior y la referida al quintil tres en la parte inferior. Sin embargo, cuando la pendiente es alta se produce un cambio y, es que, los estudiantes del quintil tres se sitúan en la parte superior del gráfico, destacando entre los estudiantes del resto de quintiles, mientras que, los situados en el quintil dos son los que muestran un menor crecimiento.

A la vista de estos resultados, parece que un aula situada en el quintil cuatro, a los estudiantes que menos beneficia es a los situados en el quintil tres, a no ser que la pendiente del aula sea alta, ya que, en este caso son los estudiantes de dicho quintil los que más crecen entre la primera ocasión de medida y la última. En cuanto a los estudiantes que se encuentran situados en el quintil cuatro, asistir a una escuela o aula, que parte de un rendimiento previo que se sitúa también en ese quintil es una buena opción, puesto que, su rendimiento en comparación con el rendimiento de los estudiantes que pertenece al resto de quintiles, no se ve perjudicado, incluso si las aulas cuentan con una pendiente de crecimiento baja o media, son los estudiantes que muestran una mejor tasa de crecimiento.

Por último, se muestra los resultados de las aulas que parten de un rendimiento promedio alto, por ello se sitúan en el quintil cinco.

Tabla VI.56. *Resultados del modelo 14 quintil 5*

MODELO 14 QUINTIL 5													
AULA	$Z(r_{0ij})\_q$	N	Y_{0ij}	Y_{8ij}	$Y_{8ij} - Y_{0ij}$	$Z(u_{00j})$	$Z(u_{10j})$	$Z(r_{0ij})$	u_{00j}	u_{10j}	r_{0ij}	e_{0ij}	e_{8ij}
386	1	7	5,48	-6,54	-12,02	1,36	**-1,19 (Baja)**	-1,61	3,16	-1,93	-1,96	-0,33	-0,1
	2	6	7,36	-5,39	-12,75			-0,48			-0,59	0,19	-0,31
	3	5	8,26	-4,27	-12,53			-0,16			-0,2	0,7	0,42
	4	5	8,46	-3,65	-12,11			0,55			0,65	0,05	0,19
	5	4	9,32	-2,62	-11,94			1,54			1,85	-0,28	0,02
169	1	4	6,66	13,38	6,72	1,36	**0,26 (Media)**	-1,41	3,16	0,51	-1,72	0,61	0,06
	2	5	7,29	14,47	7,18			-0,66			-0,81	0,33	0,24
	3	3	8,11	14,73	6,62			-0,12			-0,15	0,5	-0,16
	4	8	8,28	15,27	6,99			0,46			0,54	-0,02	-0,3
	5	5	9,07	16,34	7,27			1,1			1,32	-0,01	-0,01
62	1	2	5,81	35,76	29,95	1,37	**1,96 (Alta)**	-1,71	3,18	3,39	-2,07	0,11	-0,25
	2	4	6,9	37,9	31			-0,57			-0,69	-0,18	0,51
	3	7	8,36	37,85	29,49			-0,03			-0,05	0,63	-0,19
	4	5	8,25	38,29	30,04			0,68			0,81	-0,34	-0,6
	5	5	10,33	40,22	29,89			1,89			2,27	0,28	-0,14



Dentro de las aulas situadas en el quintil cinco, cuando los estudiantes se encuentran situados en aulas con una pendiente de crecimiento baja (-1,19), como es el aula 386, se observa que los estudiantes de los diferentes quintiles tienen una pérdida en su tasa de rendimiento, desde la primera ocasión de medida a la última, de unos 12-13 puntos. Siendo los estudiantes situados en el quintil cinco, los que menos ven disminuido su rendimiento (11,94 puntos), y los estudiantes situados en el quintil dos, los que presentan una peor tasa de crecimiento, su rendimiento disminuye unos 12,75 puntos.

Cuando los estudiantes asisten a un aula con una pendiente de crecimiento de 0,26 (pendiente media), como es el caso del aula 169, son los estudiantes situados en el quinto quintil los que más incrementan su rendimiento, ya que, mientras el resto

de estudiante obtienen una ganancia en la escala de rendimiento comprendida entre 6 y 7 puntos, los estudiantes del quintil cinco incrementan su rendimiento entre la primera ocasión de medida y la última 7,27 puntos, muy seguidos por la tasa de crecimiento de los estudiantes situados en el quintil dos (7,18 puntos). Sin embargo, tal y como también se puede apreciar en el gráfico de la tabla anterior, cuando la pendiente del aula es alta (aula 62), ya no son los estudiantes situados en el quintil cinco los que muestran una tasa de crecimiento superior, sino, los del quintil dos, pues su rendimiento se incrementa 31 puntos. Que no sean los estudiantes del quintil cinco los que obtienen una mayor ganancia puede ser debido a un efecto techo.

A la vista de estos resultados de la tabla anterior VI.56, asistir a un aula que en promedio parte de un nivel de rendimiento alto, beneficia especialmente a los estudiantes situados en el quintil cinco, aunque cuando asisten a un aula situada en este quintil, con una pendiente de crecimiento alta, es posible que se dé el conocido efecto techo. En cuanto a los estudiantes situados en el quintil uno, asistir a un aula que está formada sobre todo por estudiantes que parten de un rendimiento previo alto, no les afecta negativamente, ya que, si se compara la tasa de crecimiento de estos estudiantes con la de los estudiantes situados en el resto de quintiles, no son los peores posicionados. Es más cuando asisten a un aula con una pendiente alta, su tasa de crecimiento se incrementa casi 30 puntos. Por lo tanto, que estudiantes con bajo rendimiento de partida asistan a aulas compuestas por estudiantes que en su mayoría parten de un estatus alto, puede ser un punto a favor para los estudiantes situados en el primer quintil, ya que, su tasa de rendimiento no se ve afectada negativamente.

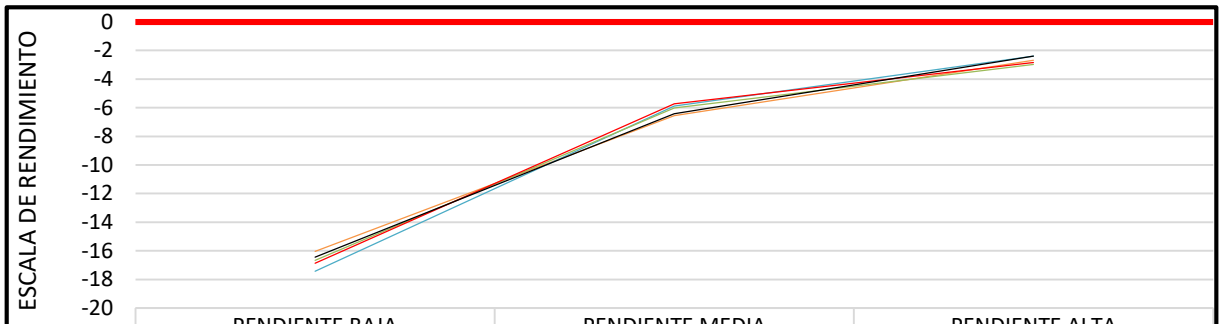
En las cinco tablas siguientes, se muestran las aulas seleccionadas para cada uno de los cinco quintiles a partir de los resultados obtenidos en el modelo 14.1. En dicho modelo se fijó una correlación entre el intercepto y la pendiente del nivel de aula alta (0,9), por lo tanto, el nivel de partida de las aulas se encuentra muy relacionado con la pendiente de crecimiento de las mismas.

En este modelo para cada uno de los quintiles, fue posible seleccionar las mismas aulas que las escogidas en el modelo 14, donde la correlación entre el intercepto y la pendiente del nivel de aula también era positiva, pero más baja (0,2).

En la siguiente tabla, se presentan los resultados de los estudiantes que asisten a aulas en las cuales en promedio parten de un rendimiento que es bajo, por ello se han situado en el primer quintil

Tabla VI.57. *Resultados del modelo 14.1 quintil 1*

MODELO 14.1 QUINTIL 1													
AULA	$Z(r_{0ij})\_q$	**N**	Y_{0ij}	Y_{8ij}	$Y_{8ij} - Y_{0ij}$	$Z(u_{00j})$	$Z(u_{10j})$	$Z(r_{0ij})$	u_{00j}	u_{10j}	r_{0ij}	e_{0ij}	e_{8ij}
266	1	5	-0,67	-16,72	-16,05	-1,4	**-1,68 (Baja)**	-1,38	-3,06	-2,52	-1,67	-0,55	0,33
	2	4	1,17	-16,25	-17,42			-0,52			-0,64	0,27	-0,23
	3	6	1,53	-15,14	-16,67			-0,05			-0,07	0,06	0,31
	4	3	2,33	-14,53	-16,86			0,59			0,7	0,08	0,15
	5	7	3,17	-13,27	-16,44			1,27			1,52	0,11	0,59
286	1	9	0	-6,55	-6,55	-1,41	**-0,81 (Media)**	-1,22	-3,08	-1,17	-1,49	-0,04	-0,4
	2	2	0,25	-5,63	-5,88			-0,72			-0,88	-0,39	-0,09
	3	5	1,16	-4,85	-6,01			0			-0,01	-0,36	-0,19
	4	4	1,85	-3,87	-5,72			0,55			0,65	-0,32	0,14
	5	5	3,37	-3,05	-6,42			1,34			1,61	0,24	0
173	1	4	0,38	-2,31	-2,69	-1,35	**-0,49 (Alta)**	-1,19	-2,94	-0,68	-1,44	0,16	-0,3
	2	11	1,23	-1,16	-2,39			-0,48			-0,59	0,16	-0,01
	3	4	1,79	-1,18	-2,97			0,11			0,12	0	-0,73
	4	1	2,3	-0,54	-2,84			0,31			0,36	0,28	-0,33
	5	7	3,4	1	-2,4			1,46			1,76	-0,02	-0,19



	PENDIENTE BAJA	PENDIENTE MEDIA	PENDIENTE ALTA
q1	-16,05	-6,55	-2,69
q2	-17,42	-5,88	-2,39
q3	-16,67	-6,01	-2,97
q4	-16,86	-5,72	-2,84
q5	-16,44	-6,42	-2,4

Lo primero que se observa en la tabla anterior VI.57, es que, los estudiantes que asisten a un aula situada en el primer quintil, tienen una pérdida en su tasa de rendimiento, entre la primera ocasión de medida y la última. Por ello, tal y como se observa en el gráfico, todas las líneas se encuentran por debajo de cero, incluso cuando los estudiantes asisten a un aula con una pendiente alta.

Si los estudiantes van a un aula con una pendiente baja (-1,68), al igual que ocurría en el modelo 14, la tasa de crecimiento de los mismos decrece entre la primera ocasión de medida y la última, pero mientras que en el modelo 14, el rendimiento de estos estudiantes disminuía entre 11 y 13 puntos, en este modelo 14.1, el decrecimiento es más marcado, entre unos 16 y 18 puntos. En ambos modelos, son los estudiantes situados en el primer quintil los que menos ven disminuido su rendimiento, quizás debido a que asisten a un aula donde la mayoría de los estudiantes parten de un nivel de rendimiento bajo. Del mismo modo, también en dichos modelos, son los estudiantes situados en el segundo quintil los más perjudicados. Concretamente en el modelo 14.1, la tasa de crecimiento de dichos estudiantes disminuye 12,67 puntos. Por lo tanto, aunque dichos estudiantes también parten de un rendimiento bajo, son los que presentan una mayor ganancia negativa al acudir a este tipo de aula.

En cuanto a los estudiantes que asisten a un aula del primer quintil, pero con una pendiente de crecimiento media (aula 286), se observa una disminución de la puntuación promedia de todos los estudiantes, desde la primera ocasión de medida a la última, comprendida entre 5 y 7 puntos. Sin embargo, en el modelo 14, los estudiantes del aula 286, no mostraban una ganancia negativa en su tasa de crecimiento, sino un crecimiento positivo superior a 11 puntos. Esto puede deberse, a que en el modelo 14, la pendiente de crecimiento del aula 286 era positiva (0,62), cosa que no ocurre en el modelo 14.1, puesto que, en las aulas de esta situación el punto de partida está muy relacionado con el crecimiento (correlación de 0,9 entre el intercepto y la pendiente) y, en este caso el residuo del estatus es negativo.

Finalmente, en cuanto a los estudiantes que asisten a un aula con una pendiente de crecimiento considerada como alta, se observa que todos ellos disminuyen su rendimiento entre 2 y 3 puntos. En esta misma aula (173), pero referida al modelo 14, donde la pendiente de crecimiento era de 1,17, se observó una situación muy diferente, en la cual, los estudiantes de todos los quintiles incrementaban su rendimiento en más de 18 puntos.

Que la tasa de crecimiento de los estudiantes que asisten a las diferentes aulas del primer quintil del modelo 14.1, disminuya entre la primera ocasión de medida y la última, es debido a que, las aulas cuentan con unos residuos de estatus muy bajos y negativos (-1,4, -1,41 y -1,35, respectivamente), que se encuentran muy

relacionados con los de la pendiente de crecimiento, ya que, se fijó una correlación entre el intercepto y la pendiente del nivel de aula de 0,9. Por ello, a la hora de resolver la ecuación final (ver ecuación V.1) , los resultados obtenidos en la última medida son negativos.

En el modelo 14, aunque también algunas aulas contaban con residuos en el intercepto negativo, los de la pendiente no siempre lo eran, ya que, no existía tanta relación entre el estatus y la pendiente, debido a que, la correlación fijada entre ambos coeficientes era baja (0,2).

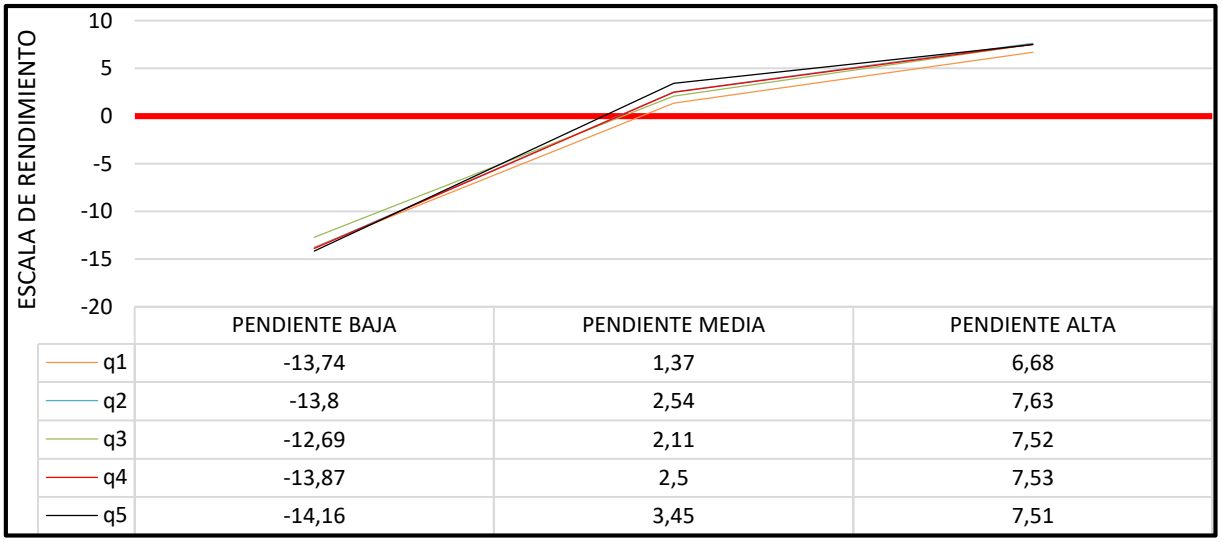
Por lo tanto, si se compara las dos tablas de resultados referidas a las aulas del primer quintil del modelo 14 y del quintil uno del modelo 14.1, se observa que cuando la correlación entre el intercepto y la pendiente es 0,9, los estudiantes muestran una ganancia negativa en su tasa de crecimiento. Mientras que en el modelo 14, donde la correlación entre el intercepto y la pendiente se fijó a 0,2, sólo los estudiantes que asisten a un aula con una pendiente baja, ven su tasa de crecimiento reducida a lo largo del tiempo, pero si asisten a un aula con una pendiente media o alta la ganancia a lo largo de las mediciones es positiva.

A la vista de todo esto, es preferible, que dentro de las aulas situadas en el primer quintil, los estudiantes asistan a aulas donde no existe mucha relación entre el rendimiento de partida y la pendiente de crecimiento, ya que, si la correlación entre ambos coeficientes es alta y de direccionalidad positiva (0,9), ningún estudiante, independientemente del quintil en el que se situé y al tipo de aula al que asista (pendiente baja, media o alta), obtendrá una ganancia positiva entre la primera ocasión de medida y la última.

A continuación, se presentan los resultados referidos a los estudiantes que asisten a aulas situadas en el quintil dos.

Tabla VI.58. *Resultados del modelo 14.1 quintil 2*

MODELO 14.1 QUINTIL 2													
AULA	$Z(r_{0ij})$ _q	**N**	Y_{0ij}	Y_{8ij}	$Y_{8ij} - Y_{0ij}$	$Z(u_{00j})$	$Z(u_{10j})$	$Z(r_{0ij})$	u_{00j}	u_{10j}	r_{0ij}	e_{0ij}	e_{8ij}
215	1	6	1,74	-12	-13,74	-0,64	**-1,42 (Baja)**	-1,52	-1,33	-2,11	-1,84	0,32	0,24
	2	6	3,01	-10,79	-13,8			-0,52			-0,64	0,38	0,24
	3	5	2,8	-9,89	-12,69			0,04			0,03	-0,5	0,47
	4	3	4,19	-9,68	-13,87			0,49			0,59	0,34	0,13
	5	2	5,09	-9,07	-14,16			1,18			1,41	0,42	-0,09
248	1	4	2,16	3,53	1,37	-0,61	**-0,26 (Media)**	-1,5	-1,27	-0,1	-1,82	0,65	-0,36
	2	5	2,7	5,24	2,54			-0,64			-0,79	0,16	0,32
	3	10	3,64	5,75	2,11			0			-0,01	0,32	0,05
	4	6	4,38	6,88	2,5			0,57			0,68	0,37	0,49
	5	3	4,37	7,82	3,45			1,29			1,55	-0,51	0,55
185	1	1	1,64	8,32	6,68	-0,59	**-0,12 (Alta)**	-1,11	-1,23	0,5	-1,35	-0,38	-0,86
	2	7	2,32	9,95	7,63			-0,53			-0,64	-0,4	0,06
	3	4	3,23	10,75	7,52			0,06			0,07	-0,2	0,15
	4	8	4,07	11,6	7,53			0,57			0,67	0,02	0,39
	5	5	5,03	12,54	7,51			1,4			1,68	-0,01	0,34



En la tabla referida a las aulas del quintil dos del modelo 14.1, a diferencia de los resultados obtenidos para las aulas del primer quintil, no todos los estudiantes obtienen una ganancia negativa, entre la primera ocasión de medida y la última, sino que depende del tipo de aula al que asistan (pendiente baja, media o alta).

Los estudiantes que acuden a un aula con una pendiente baja (aula 215), su rendimiento disminuye entre la primera ocasión de medida y el resto de ocasiones, entre 12 y 14,5 puntos. Los estudiantes situados en esta misma aula, pero del modelo 14, obtenían una pérdida de más de 21 puntos. Siendo, en ambos modelos, los estudiantes situados en el quintil tres, los que menos ganancia negativa tenían y los situados en el quintil cinco los que más.

En la tabla anterior VI.58, se observa que los estudiantes que asisten a un aula con una pendiente media, siendo la misma negativa, su rendimiento no se ve reducido entre la primera ocasión de medida y la última, sino que se incrementa alrededor de 2 puntos, excepto en los estudiantes situados en el primer quintil, que tan solo incrementan su tasa de crecimiento en 1,37 puntos. Esta misma situación, pero referido al modelo 14, donde la correlación entre el intercepto y la pendiente del nivel de aula se fijó a 0,2, los estudiantes de los distintos quintiles tenían una ganancia en rendimiento de más de 13 puntos.

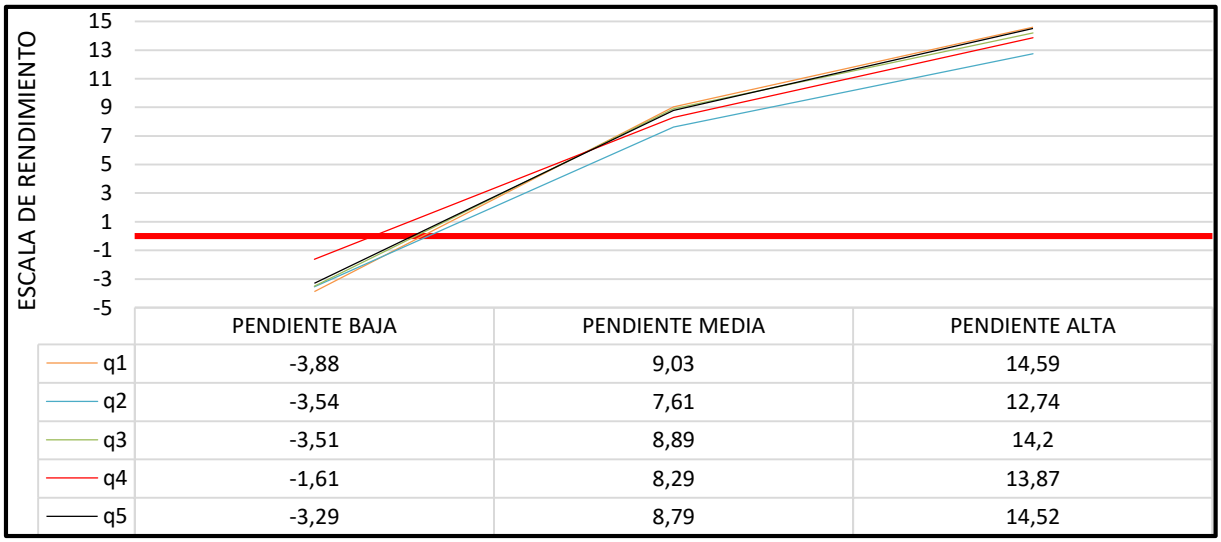
Finalmente, se observa que los estudiantes que asisten a un aula con una pendiente alta, obtienen una ganancia comprendida entre 6 y 8 puntos, mientras que, en el modelo 14 dicha ganancia era superior a 23 puntos en la escala de rendimiento.

Por lo tanto, a la vista de los resultados anteriores, si los estudiantes tienen que acudir a aulas situadas en el quintil dos, pero con una pendiente baja, es mejor que asistan a aulas donde el rendimiento de partida se encuentra muy relacionado con la pendiente de crecimiento, ya que, la reducción en la tasa de rendimiento que se produce en estos estudiantes, es menor que la obtenida en el modelo 14. Sin embargo, si los estudiantes asisten a un aula con pendiente media o alta, es más adecuado que asistan a aulas donde la correlación entre el intercepto y la pendiente es baja (0,2), que cuando es alta (0,9), ya que, la tasa de crecimiento en este último modelo es notablemente más baja.

Posteriormente, en la tabla siguiente, se muestran los resultados de las tres aulas seleccionadas dentro del quintil tres.

Tabla VI.59. *Resultados del modelo 14.1 quintil 3*

MODELO 14.1 QUINTIL 3													
AULA	$Z(r_{0ij})$ _q	N	Y_{0ij}	Y_{8ij}	$Y_{8ij} - Y_{0ij}$	$Z(u_{00j})$	$Z(u_{10j})$	$Z(r_{0ij})$	u_{00j}	u_{10j}	r_{0ij}	e_{0ij}	e_{8ij}
90	1	5	3,88	0	-3,88	0,09	**-0,58 (Baja)**	-1,27	0,3	-0,81	-1,55	0,53	-0,11
	2	7	4,29	0,75	-3,54			-0,57			-0,7	0,09	-0,2
	3	7	5,02	1,51	-3,51			0,05			0,05	0,07	-0,2
	4	2	4,71	3,1	-1,61			0,46			0,55	-0,74	0,9
	5	3	6,67	3,38	-3,29			1,35			1,62	0,15	0,11
384	1	6	3,45	12,48	9,03	0,12	**0,4 (Media)**	-1,31	0,37	0,7	-1,59	0,07	0,29
	2	3	4,9	12,51	7,61			-0,51			-0,62	0,55	-0,65
	3	5	5,18	14,07	8,89			0,01			0	0,21	0,28
	4	8	5,89	14,18	8,29			0,55			0,66	0,26	-0,26
	5	4	6,04	14,83	8,79			1,14			1,37	-0,3	-0,32
261	1	3	2,08	16,67	14,59	0,14	**0,81 (Alta)**	-2,02	0,41	1,34	-2,45	-0,49	0,19
	2	6	4,82	17,56	12,74			-0,56			-0,69	0,49	-0,67
	3	8	5,09	19,29	14,2			0,02			0,02	0,06	0,35
	4	2	5,91	19,78	13,87			0,58			0,69	0,21	0,17
	5	8	6,36	20,88	14,52			1,25			1,5	-0,15	0,46



Si los resultados de la tabla anterior (Tabla VI.59), se comparan con los resultados de la Tabla VI.54, referida a las aulas del quintil tres del modelo 14, se observa que cuando los estudiantes acuden a un aula con una pendiente baja, en el modelo 14.1, los estudiantes disminuyen su rendimiento entre 3 y 4 puntos, a excepción de los estudiantes del quintil cuatro, que tan sólo lo disminuyen 1,61 puntos, por eso en el gráfico dicha línea se encuentra muy cerca de la línea roja que representa el 0. Sin embargo, en el modelo 14, aunque también los estudiantes disminuían su rendimiento entre la primera ocasión y la última, dicha disminución era superior a 14 puntos en los estudiantes de los diferentes quintiles, excepto los del quintil cuatro, que su tasa de crecimiento se veía disminuida 12,49 puntos.

Cuando los estudiantes asisten al aula 384, en el modelo donde el estatus esta muy relacionado con la pendiente de crecimiento, incrementan su rendimiento más de 8 puntos, menos los estudiantes situados en el quintil dos, cuyo incremento es de 7,61. En el aula 384 del modelo 14, también eran los estudiantes situados en el quintil dos, los que presentaban una menor tasa de crecimiento, ya que, su rendimiento se incrementaba entre la primera ocasión de medida y la última unos 11 puntos, y en el resto de estudiantes era superior a 11,5 puntos.

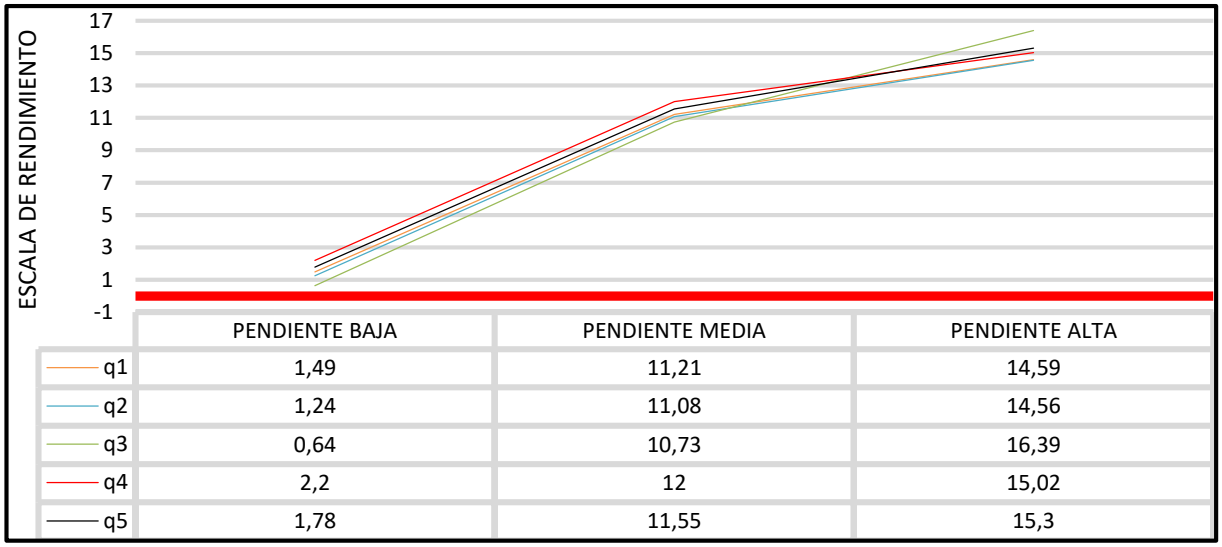
Finalmente, en cuanto, a los estudiantes que asisten al aula 261 del modelo 14.1, incrementan su rendimiento entre 12 y 14,5 puntos, mientras que, en el modelo 14 la mayoría de los estudiantes situados en los diferentes quintiles, su tasa de crecimiento aumentaba más de 23 puntos.

A la vista de estos resultados, se llega a la misma conclusión que en las aulas del quintil dos, y, es que, si los estudiantes asisten a aulas con una pendiente baja, tienen una mayor pérdida en el modelo 14, pero si asisten a aulas con una pendiente media o alta, su tasa de crecimiento es mayor en el modelo 14, respecto al modelo14.1.

A continuación, se muestran los resultados de las aulas seleccionadas dentro del cuarto quintil.

Tabla VI.60. *Resultados del modelo 14.1 quintil 4*

MODELO 14.1 QUINTIL 4													
AULA	$Z(r_{0ij})$ _q	N	Y_{0ij}	Y_{8ij}	$Y_{8ij} - Y_{0ij}$	$Z(u_{00j})$	$Z(u_{10j})$	$Z(r_{0ij})$	u_{00j}	u_{10j}	r_{0ij}	e_{0ij}	e_{8ij}
14	1	8	4,2	5,69	1,49	0,56	**-0,25 (Baja)**	-1,26	1,36	-0,3	-1,54	-0,23	0,47
	2	2	5,18	6,42	1,24			-0,62			-0,76	-0,02	0,42
	3	9	6,11	6,75	0,64			0,05			0,05	0,1	-0,05
	4	3	5,74	7,94	2,2			0,59			0,7	-0,91	0,49
	5	3	6,63	8,41	1,78			1,03			1,24	-0,57	0,41
113	1	8	4,71	15,92	11,21	0,58	**0,58 (Media)**	-1,15	1,41	0,99	-1,4	0,1	0,21
	2	6	5,36	16,44	11,08			-0,51			-0,62	-0,03	-0,04
	3	4	6,14	16,87	10,73			0,09			0,11	0,03	-0,34
	4	2	6,17	18,17	12			0,56			0,66	-0,5	0,4
	5	3	7,05	18,6	11,55			1,32			1,58	-0,54	-0,09
178	1	8	4,74	19,33	14,59	0,59	**0,88 (Alta)**	-1,29	1,44	1,46	-1,57	0,26	-0,01
	2	1	5,51	20,07	14,56			-0,64			-0,78	0,25	-0,07
	3	1	3,49	19,88	16,39			-0,2			-0,26	-2,29	-0,78
	4	6	6,86	21,88	15,02			0,51			0,61	0,21	0,35
	5	9	7,63	22,93	15,3			1,44			1,73	-0,13	0,29



A la vista de los resultados de la tabla anterior VI.60, se observa que, por primera vez, todos los estudiantes incrementan su tasa de rendimiento independientemente del tipo de aula al que asistan. Los estudiantes que asisten a un aula con una pendiente baja, incrementan su rendimiento entre 1 y 2 puntos, excepto los estudiantes situados en el quintil tres que no llegan ni a un punto. En el modelo 14, los estudiantes que asistían a aulas con una pendiente baja, su tasa de crecimiento decrecía alrededor de unos 14-16 puntos.

Al comparar los resultados, del modelo 14 con el modelo 14.1, los estudiantes situados en un aula con una pendiente media, no obtienen una mayor tasa de crecimiento cuando acuden a un aula donde el intercepto y la pendiente no se encuentran muy relacionados, sino, que su crecimiento es mayor, cuando el

rendimiento de partida de las aulas se encuentra muy relacionada con su tasa de crecimiento. En este caso los estudiantes incrementan su rendimiento entre 10 y 12 puntos, mientras que en el modelo 14 dicha tasa se incrementaba entre 6 y 8 puntos.

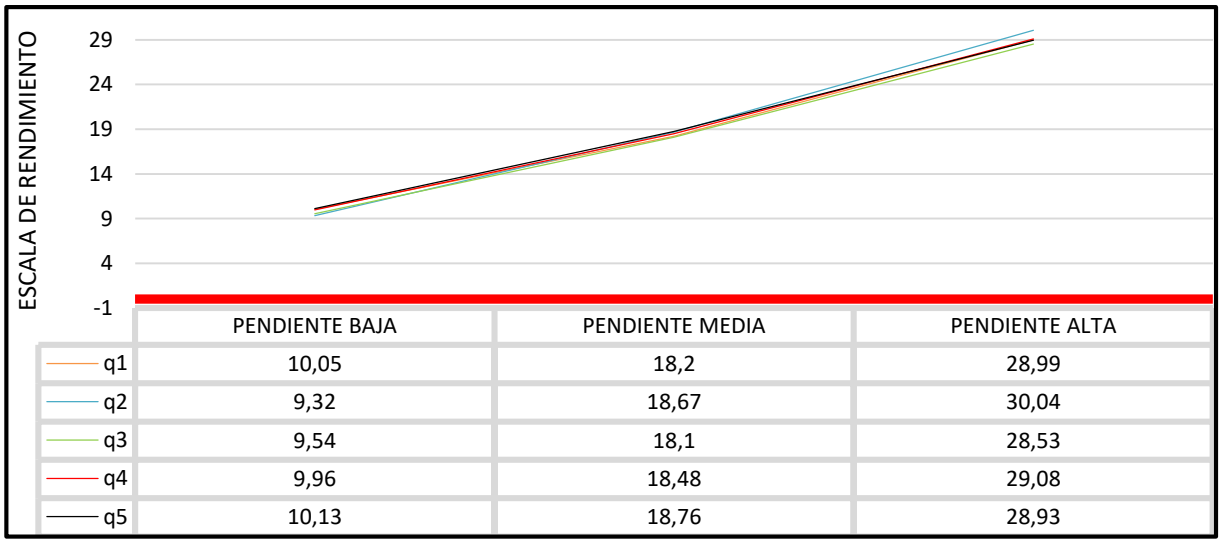
No existe tanta diferencia en la tasa de crecimiento de los estudiantes que asisten al aula 178 en el modelo 14 y en el modelo 14.1, como en las situaciones anteriores, ya que, ambos modelos muestran que los estudiantes incrementan su rendimiento más de 14 puntos entre la primera ocasión de medida y la última. Incluso dicha tasa es ligeramente superior en los estudiantes de los diferentes quintiles del modelo 14.

A la vista de los resultados de la tabla anterior (Tabla VI. 60), si se comparan los mismos, con los resultados de la Tabla VI. 55 referida a las aulas seleccionadas dentro del quintil cuatro del modelo 14, se puede concluir que, es preferible que los estudiantes que asisten a aulas situadas en el quintil cuatro, vayan a aulas donde el rendimiento de partida esta muy relacionado con la tasa de crecimiento, especialmente si el residuo de la pendiente es bajo o medio, ya que, van obtener una mayor ganancia que cuando asisten a aulas del quintil cuatro, donde el rendimiento de partida no se encuentra muy relacionado con el adquirido a lo largo del tiempo. Sin embargo, si el residuo de la pendiente es alto, aunque no se han encontrado diferencias muy marcadas entre las tasas de crecimiento obtenidas en los estudiantes del modelo 14.1 y del modelo 14, son ligeramente superiores las de este último modelo.

Finalmente, dentro del modelo 14.1 se presentan los resultados de las aulas seleccionadas dentro del quintil cinco.

Tabla VI.61. *Resultados del modelo 14.1 quintil 5*

MODELO 14.1 QUINTIL 5													
AULA	$Z(r_{0ij})\_q$	N	Y_{0ij}	Y_{8ij}	$Y_{8ij} - Y_{0ij}$	$Z(u_{00j})$	$Z(u_{10j})$	$Z(r_{0ij})$	u_{00j}	u_{10j}	r_{0ij}	e_{0ij}	e_{8ij}
386	1	7	5,43	15,48	10,05	1,33	**0,48 (Baja)**	-1,61	3,11	0,83	-1,96	-0,33	-0,1
	2	6	7,31	16,63	9,32			-0,48			-0,59	0,19	-0,31
	3	5	8,21	17,75	9,54			-0,16			-0,2	0,7	0,42
	4	5	8,41	18,37	9,96			0,55			0,65	0,05	0,19
	5	4	9,27	19,4	10,13			1,54			1,85	-0,28	0,02
169	1	4	6,66	24,86	18,2	1,36	**1,2 (Media)**	-1,41	3,16	1,94	-1,72	0,61	0,06
	2	5	7,29	25,96	18,67			-0,66			-0,81	0,33	0,24
	3	3	8,11	26,21	18,1			-0,12			-0,15	0,5	-0,16
	4	8	8,28	26,76	18,48			0,46			0,54	-0,02	-0,3
	5	5	9,07	27,83	18,76			1,1			1,32	-0,01	-0,01
62	1	2	5,87	34,86	28,99	1,39	**2,05 (Alta)**	-1,71	3,24	3,27	-2,07	0,11	-0,25
	2	4	6,96	37	30,04			-0,57			-0,69	-0,18	0,51
	3	7	8,42	36,95	28,53			-0,03			-0,05	0,63	-0,19
	4	5	8,31	37,39	29,08			0,68			0,81	-0,34	-0,6
	5	5	10,39	39,32	28,93			1,89			2,27	0,28	-0,14



En la tabla anterior del quintil cinco, se observa algo similar a lo que ocurría en la Tabla VI.60 referida a las aulas del quintil cuatro, y es que, todos los estudiantes incrementan su tasa de rendimiento desde la primera ocasión de medida a la última, independientemente del valor de la pendiente y del quintil en el que se sitúen. Sin embargo, en la tabla del quintil cinco del modelo 14, los estudiantes situados en un aula con una pendiente baja, sufrían una disminución en su tasa de crecimiento desde la primera ocasión de medida a la última. Mientras que en la tabla anterior los estudiantes incrementan su rendimiento en más de 9 puntos, siendo los estudiantes situados en el último quintil y en el primer quintil, los que más incrementan su rendimiento. Por lo tanto, a los estudiantes que parten de un rendimiento bajo, les

beneficia acudir a aulas compuestas sobre todo por estudiantes con alto rendimiento de partida.

En cuanto a los estudiantes situados en un aula donde la pendiente es media, ven incrementado su rendimiento más de 18 puntos, tal y como, se puede observar en la tabla anterior VI.61. Sin embargo, en el modelo 14, el incremento de la tasa de crecimiento de estos estudiantes era alrededor de unos 6 o 7 puntos.

Finalmente, la ganancia producida en los estudiantes que acuden a un aula con un valor en el residuo de la pendiente alto, se observa que en ambos modelos (modelo 14 y modelo 14.1), los estudiantes obtienen una ganancia en crecimiento superior a 29 puntos, siendo esta ganancia ligeramente superior en el modelo 14.

A la vista de todos estos resultados, al comparar las aulas de los distintos quintiles del modelo 14, con los del modelo 14.1, en determinadas situaciones se llegan a conclusiones muy similares acerca del rendimiento de los estudiantes de un determinado quintil, en cuanto a los estudiantes que más y menos incrementan su tasa de crecimiento. Aunque el incremento de la misma, por lo general, es bastante distinta, debido al efecto de la correlación entre el intercepto y la pendiente del nivel de aula.

En las aulas situadas en el primer quintil con una pendiente baja, media o alta (aulas 266, 286 y 173) los estudiantes de todos los quintiles muestran una mejor tasa de rendimiento cuando la correlación entre el intercepto y la pendiente es baja (0,2). Sin embargo, en las aulas del quintil dos y tres con una pendiente baja, los estudiantes sufren una mayor pérdida en su tasa de crecimiento en el modelo 14.1, donde el rendimiento de partida se encuentra muy relacionado con el adquirido a lo largo del tiempo, pero los estudiantes que asisten a aulas con una pendiente media y alta, rinden más cuando los niveles de logro en la primera ocasión de medida no se encuentran muy relacionado con la pendiente de crecimiento (modelo 14). Respecto a las aulas situadas en el quintil cuatro y cinco, si su pendiente de crecimiento es baja o media, los estudiantes de los diferentes quintiles obtienen una mayor ganancia en el modelo 14.1, pero si la pendiente de crecimiento de las aulas es alta, obtienen resultados muy similares tanto cuando el rendimiento de partida se encuentra muy relacionado con el adquirido a lo largo del tiempo (modelo 14.1), como cuando dicha

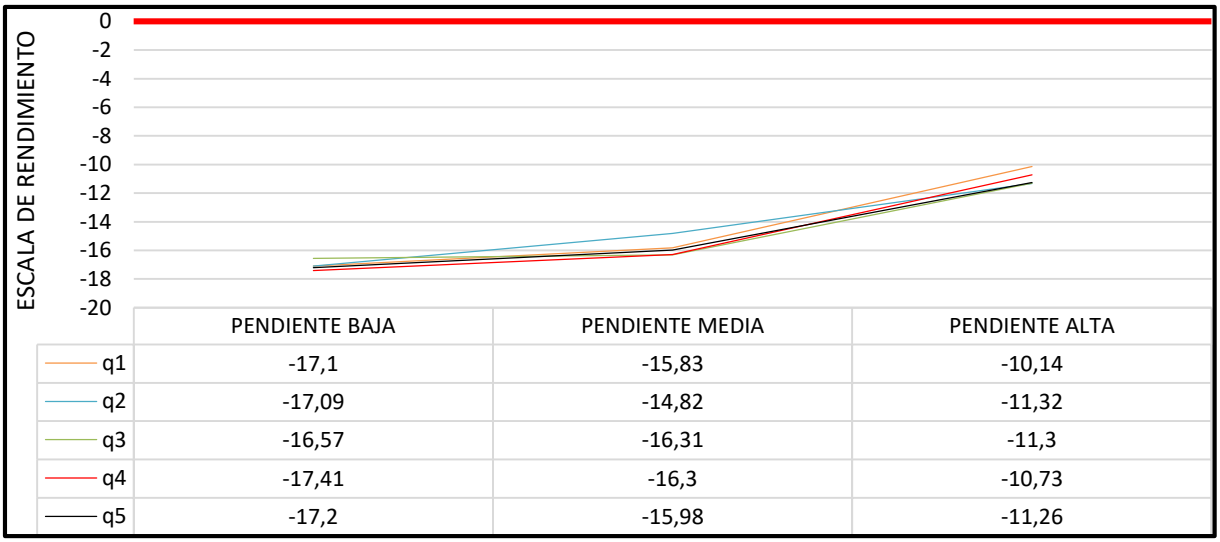
relación no es tan pronunciada (modelo 14), incluso la tasa de crecimiento es ligeramente superior en esta situación.

Si se pudiera elegir las aulas de un quintil que no perjudicaran al rendimiento de ningún estudiante y potenciara el mismo, independientemente del valor del residuo de la pendiente, sería el quintil cinco del modelo 14.1, ya que, es cuando los estudiantes muestran una mayor tasa de crecimiento. Si en vez de un quintil, hubiera que seleccionar un aula, el que más ganancia produce en la tasa de rendimiento de los estudiantes es el aula 62, correspondiente al quintil cinco, tanto del modelo 14 como del 14.1, ya que, los estudiantes de los diferentes quintiles incrementan su tasa de rendimiento en más de 28 puntos.

A continuación, se presentan los resultados de los cinco quintiles en el modelo 14.2, en el cual, se fijó una correlación entre el intercepto y la pendiente del nivel de aula baja-negativa (-0,2). Concretamente, en la siguiente Tabla VI.62 se presenta los resultados de los estudiantes que asisten a aulas del primer quintil.

Tabla VI.62. *Resultados del modelo 14.2 quintil 1*

MODELO 14.2 QUINTIL 1													
AULA	$Z(r_{0ij})$ _q	N	Y_{0ij}	Y_{8ij}	$Y_{8ij} - Y_{0ij}$	$Z(u_{00j})$	$Z(u_{10j})$	$Z(r_{0ij})$	u_{00j}	u_{10j}	r_{0ij}	e_{0ij}	e_{8ij}
240	1	5	0,75	-16,35	-17,1	-1,48	**-1,09 (Baja)**	-1,41	-3,23	-1,79	-1,71	0,02	-0,08
	2	4	1,71	-15,38	-17,09			-0,66			-0,8	0,07	-0,01
	3	4	2,44	-14,13	-16,57			0,09			0,1	-0,11	0,34
	4	5	3,08	-14,33	-17,41			0,48			0,57	0,07	-0,33
	5	8	3,97	-13,23	-17,2			1,36			1,64	-0,11	-0,3
67	1	3	0,05	-15,78	-15,83	-1,46	**-0,44 (Media)**	-1,97	-3,19	-0,72	-2,39	0,41	-0,06
	2	7	0,95	-13,87	-14,82			-0,64			-0,78	-0,29	0,25
	3	5	2,34	-13,97	-16,31			-0,04			-0,06	0,38	-0,58
	4	5	3,32	-12,98	-16,3			0,57			0,68	0,62	-0,33
	5	6	3,84	-12,14	-15,98			1,44			1,73	0,1	-0,54
188	1	2	-2,58	-12,72	-10,14	-1,42	**1,03 (Alta)**	-2,28	-3,1	1,74	-2,76	-0,88	0,5
	2	8	0,17	-11,15	-11,32			-0,42			-0,52	-0,37	-0,17
	3	8	0,82	-10,48	-11,3			0,04			0,04	-0,28	-0,05
	4	4	1,22	-9,51	-10,73			0,62			0,74	-0,59	0,21
	5	3	2,6	-8,66	-11,26			1,37			1,64	-0,1	0,16



En la tabla anterior VI.62, lo primero que se observa es que todos los estudiantes ven disminuido su rendimiento desde la primera ocasión de medida a la última, por ello, en el gráfico de esta tabla todas las líneas referentes a los distintos residuos de las pendientes, se encuentran muy por debajo de la línea roja que representa el cero. Sin embargo, cuando se fija el mismo valor de correlación que en el modelo14.2, pero con direccionalidad positiva (modelo 14), tan sólo los estudiantes que asistían a un aula con un residuo en la pendiente bajo, sufrían una disminución de rendimiento entre la ocasión una y la nueve. El resto de estudiantes asistentes a aulas con un residuo de la pendiente medio o alto, incrementaban su tasa de rendimiento.

También en la tabla se observa que los estudiantes que asisten a un aula con un residuo en la pendiente baja (aula 240), disminuyen su rendimiento alrededor de unos 16, 17,5 puntos, siendo los estudiantes situados en el tercer quintil los que menos disminuyen su tasa de logro. Por el contrario, son los estudiantes situados en el quintil cuatro, los que más se ven afectado por esa disminución del rendimiento, ya que, su tasa de crecimiento entre la primera ocasión de medida y la última, disminuye 17,41 puntos en la escala.

Cuando los estudiantes asisten a un aula con un residuo en la pendiente de crecimiento medio (aula 67), los estudiantes situados en el quintil tres y cuatro son los que experimentan una pérdida mayor, de unos 16,3 puntos. Sin embargo, en esta misma aula, los estudiantes menos perjudicados, son los que se sitúan en el quintil dos, ya que, su tasa de crecimiento disminuye 14,82 puntos.

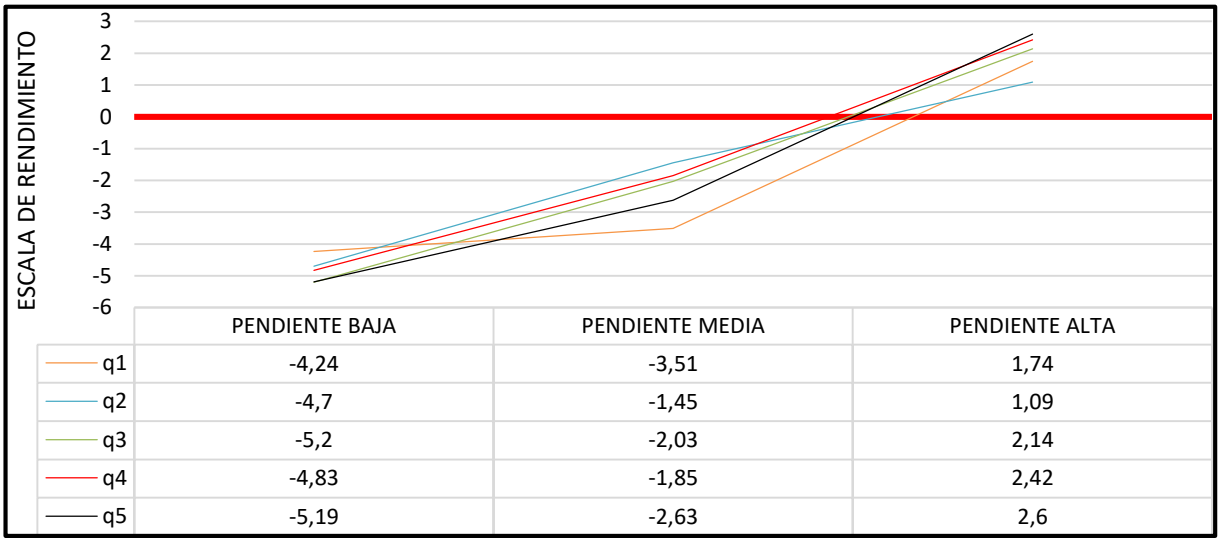
Finalmente, en cuanto a los estudiantes que asisten a un aula con un residuo en la pendiente alto (aula 188), disminuyen su rendimiento alrededor de 10 o 11,5 puntos, siendo los estudiantes situados en el primer quintil los que menos disminuyen su logro (10,14 puntos) y, los situados en el quintil dos y tres los que más (11,3 puntos).

A la vista de los resultados anteriores y, si se observa el gráfico se puede apreciar que, cuando los estudiantes pertenecen al cuarto quintil y asisten a aulas con un residuo de la pendiente bajo o medio, son los que presentan una mayor pérdida en su tasa de crecimiento, por ello se encuentran en la parte inferior del gráfico. Sin embargo, cuando asisten a un aula con un residuo alto en la pendiente, no son los estudiantes que más pérdida de ganancia tienen. Los estudiantes del quintil tres, tienen una menor reducción de su rendimiento cuando asisten a un aula con un residuo en la pendiente bajo y, una mayor reducción de su tasa de crecimiento cuando asisten a aulas con un residuo en la pendiente, medio/alto.

En la tabla siguiente VI. 63, se muestran los resultados de los estudiantes que se encuentran en aulas situadas en el segundo quintil dentro del modelo 14.2

Tabla VI.63. *Resultados del modelo 14.2 quintil 2*

MODELO 14.2 QUINTIL 2													
AULA	$Z(r_{0ij})$ _q	N	Y_{0ij}	Y_{8ij}	$Y_{8ij} - Y_{0ij}$	$Z(u_{00j})$	$Z(u_{10j})$	$Z(r_{0ij})$	u_{00j}	u_{10j}	r_{0ij}	e_{0ij}	e_{8ij}
139	1	9	2,44	-1,8	-4,24	-0,53	**-0,87 (Baja)**	-1,31	-1,08	-1,43	-1,59	-0,2	0,22
	2	3	3,45	-1,25	-4,7			-0,42			-0,52	-0,26	-0,29
	3	5	4,61	-0,59	-5,2			0			-0,01	0,39	-0,14
	4	8	4,6	-0,23	-4,83			0,52			0,62	-0,26	-0,42
	5	3	6,87	1,68	-5,19			1,47			1,77	0,87	0,34
198	1	7	2,87	-0,64	-3,51	-0,48	**-0,07 (Media)**	-1,1	-0,97	-0,09	-1,33	0,45	-0,78
	2	6	3,12	1,67	-1,45			-0,41			-0,5	-0,14	0,7
	3	5	3,76	1,73	-2,03			0,01			0	0	0,26
	4	4	3,99	2,14	-1,85			0,54			0,65	-0,41	0,02
	5	4	5,85	3,22	-2,63			1,19			1,43	0,67	0,32
102	1	6	0,81	2,55	1,74	-0,46	**1,64 (Alta)**	-1,61	-0,92	2,76	-1,96	0,18	0,05
	2	4	1,82	2,91	1,09			-0,52			-0,63	-0,14	-0,92
	3	3	2,5	4,64	2,14			0,07			0,07	-0,17	0,11
	4	9	2,73	5,15	2,42			0,46			0,54	-0,41	0,15
	5	2	3,6	6,2	2,6			1,07			1,28	-0,27	0,47



Lo primero que se observa en la tabla VI.63, es que, ya no todos los estudiantes asistentes a aulas del segundo quintil ven reducido su rendimiento, pues los estudiantes del aula 102, que cuentan con un residuo en su pendiente alto, experimentan un crecimiento en su tasa de logro entre 1 punto y 2,5 puntos, desde la primera ocasión de medida a la última. Sin embargo, en el modelo 14, los estudiantes pertenecientes a aulas del quintil dos, sólo sufrían un decrecimiento en su rendimiento cuando asistían a un aula con un residuo bajo en la pendiente. En el resto de aulas su tasa de crecimiento era, superior a 12 puntos cuando asistian a un aula con un residuo en la pendiente medio, y, superior a 23 puntos cuando asistían a un aula con un residuo en la pendiente alto.

También en la tabla se muestra, que cuando los estudiantes asisten a un aula con una pendiente baja, como es el aula 139, son los estudiantes situados en el quintil tres, muy seguidos por los del quintil cinco, los que más se encuentran afectados por la disminución de su tasa de rendimiento, por eso, ambas líneas están muy solapadas. En esta misma aula, son los estudiantes situados en el quintil uno los que menos se encuentran afectados por este detrimiento de la pendiente de crecimiento, ya que, su tasa de rendimiento se reduce 4,24, muy seguidos, se encuentran los estudiantes del segundo quintil que sufren una reducción en su rendimiento de 4,7 puntos. Los estudiantes situados en este mismo quintil (segundo), pero en el aula 198, donde el residuo de la pendiente es medio, son los que presentan una menor reducción de su tasa de crecimiento, ya que, la misma se ve reducida 1,45, siendo los estudiantes del quintil uno los que más reducen su logro, alrededor de unos 3,5 puntos.

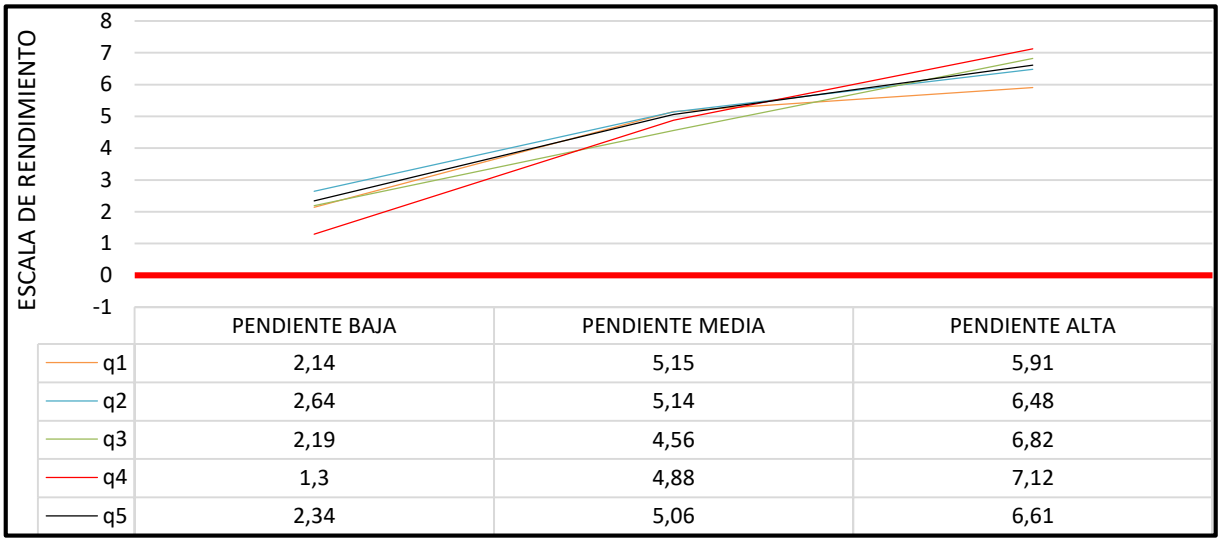
Asimismo, son los estudiantes situados en el quintil dos, junto con los del quintil uno, los que menos incrementan su tasa de crecimiento cuando asisten al aula 102, ya que, no sobrepasan los dos puntos, y son los estudiantes del quintil cinco, los que en promedio más crecen, 2,6 puntos.

Por lo tanto, a la vista de los resultados anteriores, y tal y como se puede observar en el gráfico, los estudiantes del primer quintil son los menos perjudicados cuando asisten a un aula tipo 139, pero de los menos beneficiados, cuando asisten a un aula con un residuo en la pendiente, medio-alto.

En la siguiente Tabla VI.64, se encuentra los resultados de los estudiantes pertenecientes a las aulas situadas en el tercer quintil.

Tabla VI.64. *Resultados del modelo 14.2 quintil 3*

MODELO 14.2 QUINTIL 3													
AULA	$Z(r_{0ij})\_q$	N	Y_{0ij}	Y_{8ij}	$Y_{8ij}-Y_{0ij}$	$Z(u_{00j})$	$Z(u_{10j})$	$Z(r_{0ij})$	u_{00j}	u_{10j}	r_{0ij}	e_{0ij}	e_{8ij}
55	1	7	4,14	6,28	2,14	0,03	**-0,92 (Baja)**	-1,50	0,19	-1,52	-1,82	,54	,48
	2	5	4,24	6,88	2,64			-,57			-,70	-,48	-,04
	3	6	5,23	7,42	2,19			-,08			-,11	-,08	-,09
	4	3	7	8,3	1,3			,79			,94	,65	-,26
	5	4	7,2	9,54	2,34			1,72			2,07	-,29	-,15
297	1	5	2,22	7,37	5,15	0,05	**0,22 (Media)**	-1,94	0,23	0,39	-2,35	-0,07	0,07
	2	4	4,03	9,17	5,14			-0,44			-0,54	-0,08	0,05
	3	3	4,95	9,51	4,56			0,03			0,03	0,28	-0,18
	4	9	5,34	10,22	4,88			0,49			0,58	0,11	-0,02
	5	7	6,09	11,15	5,06			1,43			1,72	-0,28	-0,23
192	1	5	2,96	8,87	5,91	0,07	**0,92 (Alta)**	-1,41	0,28	1,55	-1,71	0,47	-0,47
	2	4	3,73	10,21	6,48			-0,5			-0,62	0,15	-0,22
	3	3	3,82	10,64	6,82			0,03			0,02	-0,4	-0,44
	4	9	4,79	11,91	7,12			0,63			0,75	-0,16	0,1
	5	7	6,11	12,72	6,61			1,52			1,83	0,09	-0,16



Lo primero que se observa en el modelo 14.2 del quintil tres, es que todos los estudiantes en promedio, ven incrementado su rendimiento desde la primera ocasión de medida a la última. Cosa que no ocurría en las distintas aulas del quintil tres del modelo 14, ya que, los estudiantes que asistían a un aula con un residuo en la pendiente bajo, su rendimiento decrecía entre la primera ocasión de medida y la última.

En los resultados del modelo 14.2 (quintil 3) se aprecia que, cuando los estudiantes acuden a un aula con un residuo en la pendiente de crecimiento bajo, como es el aula 55, todos tienen una ganancia en la escala de rendimiento entre 2 y

2,7 puntos, a excepción de los estudiantes situados en el quintil cuatro, que su rendimiento tan solo crece 1,3 puntos.

Cuando el residuo de la pendiente de crecimiento del aula es medio, aula 297, los estudiantes incrementan su rendimiento entre 4 y 5 puntos, siendo los estudiantes situados en el tercer quintil los que menos incrementan su tasa de crecimiento (4,56) y, los del quintil uno, muy seguidos por los del quintil dos, los que más incrementan su tasa de rendimiento (5,15 y 5,14, respectivamente).

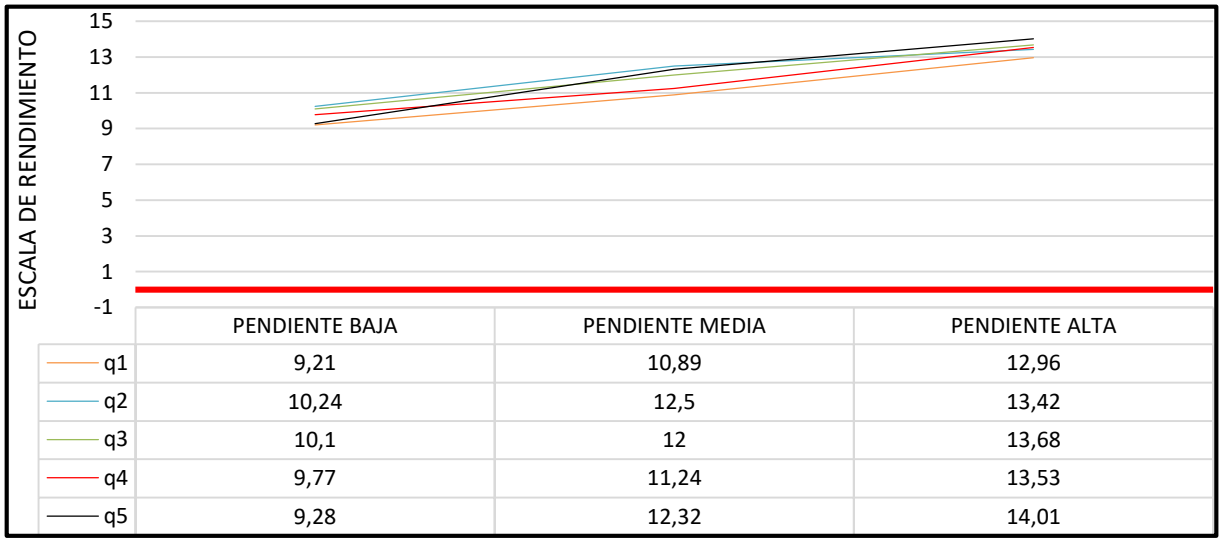
Finalmente, en el aula 192, donde el residuo de la pendiente es alto, se observa que todos los estudiantes incrementan su rendimiento más de 6 puntos, a excepción de los situados en el primer quintil, donde su ganancia es ligeramente inferior (5,91).

Tal y como se puede observar en el gráfico, los estudiantes situados en el quintil dos y en el quintil cuatro, son los que presentan un mayor cambio en su tasa de crecimiento dependiendo del aula al que asísten. Así, mientras que los estudiantes del quintil dos cuando asísten a un aula con un residuo en la pendiente bajo y medio, son los que más incrementan su tasa de rendimiento respecto a los estudiantes situados en el resto de quintiles, cuando acuden a un aula con un residuo en la pendiente bajo, son de los que menos incrementan su tasa de crecimiento. Por el contrario, los estudiantes situados en el quintil cuatro son los que presentan una menor tasa de crecimiento cuando asisten a aulas tipo 55 y la 297, pero cuando asisten a un aula con un residuo en la pendiente alto (192), son los estudiantes que en promedio más crecen entre la primera ocasión de medida y la última.

En la Tabla VI.65 que se presenta a continuación, se encuentran los resultados de los estudiantes situados en las tres aulas seleccionadas dentro del quintil cuatro.

Tabla VI.65. *Resultados del modelo 14.2 quintil 4*

MODELO 14.2 QUINTIL 4													
AULA	$Z(r_{0ij})$ _q	N	Y_{0ij}	Y_{8ij}	$Y_{8ij} - Y_{0ij}$	$Z(u_{00j})$	$Z(u_{10j})$	$Z(r_{0ij})$	u_{00j}	u_{10j}	r_{0ij}	e_{0ij}	e_{8ij}
88	1	3	4,66	13,87	9,21	0,51	-0,15 (Baja)	-1,58	1,28	-0,23	-1,92	0,73	0,03
	2	7	5,07	15,31	10,24			-0,56			-0,69	-0,1	0,24
	3	5	5,78	15,88	10,1			0,02			0,01	-0,08	0,1
	4	6	6,16	15,93	9,77			0,51			0,6	-0,29	-0,43
	5	6	7,79	17,07	9,28			1,39			1,67	0,27	-0,36
292	1	5	3,98	14,87	10,89	0,52	0,61 (Media)	-1,4	1,31	1,05	-1,71	0,35	-0,6
	2	5	4,38	16,88	12,5			-0,54			-0,66	-0,3	0,37
	3	3	5,57	17,57	12			0,08			0,09	0,14	0,31
	4	4	5,82	17,06	11,24			0,46			0,55	-0,07	-0,66
	5	6	6,61	18,93	12,32			1,39			1,67	-0,4	0,09
218	1	3	4,01	16,97	12,96	0,54	1,24 (Alta)	-1,1	1,35	2,08	-1,34	0,43	-0,04
	2	5	4,37	17,79	13,42			-0,56			-0,68	0,12	0,11
	3	4	5,07	18,75	13,68			0,07			0,08	0,07	0,32
	4	7	5,41	18,94	13,53			0,55			0,66	-0,18	-0,08
	5	6	6,75	20,76	14,01			1,76			2,11	-0,29	0,29



Como en la tabla del quintil tres (Tabla VI.64), todos los estudiantes que asisten a un aula del quintil cuatro incrementan su rendimiento entre la primera ocasión de medida y la última. Esto no ocurría en el modelo 14, en los estudiantes que asistían a un aula del quintil cuatro, con un residuo en la pendiente bajo, pues en esta situación todos los estudiantes disminuían su rendimiento más de 14 puntos.

Se observa que los estudiantes que acuden al aula 88, donde el residuo de la pendiente de crecimiento es bajo, incrementan su tasa de crecimiento en torno a 9-10 puntos, siendo los estudiantes situados en el primer quintil, seguidos de los del quinto quintil, los que muestran una menor tasa de crecimiento. Por el contrario, son

los estudiantes situados en el segundo quintil, seguidos por los del tercer quintil, los que muestran mayor crecimiento (10,24 y 10,1 puntos, respectivamente).

Esta tendencia también se mantiene en el aula 292, ya que, son los estudiantes situados en el primer quintil los que menos incrementan su tasa de crecimiento (10,89 puntos), y los situados en el quintil dos los que más crecen entre la primera ocasión de medida y la última (12,5 puntos).

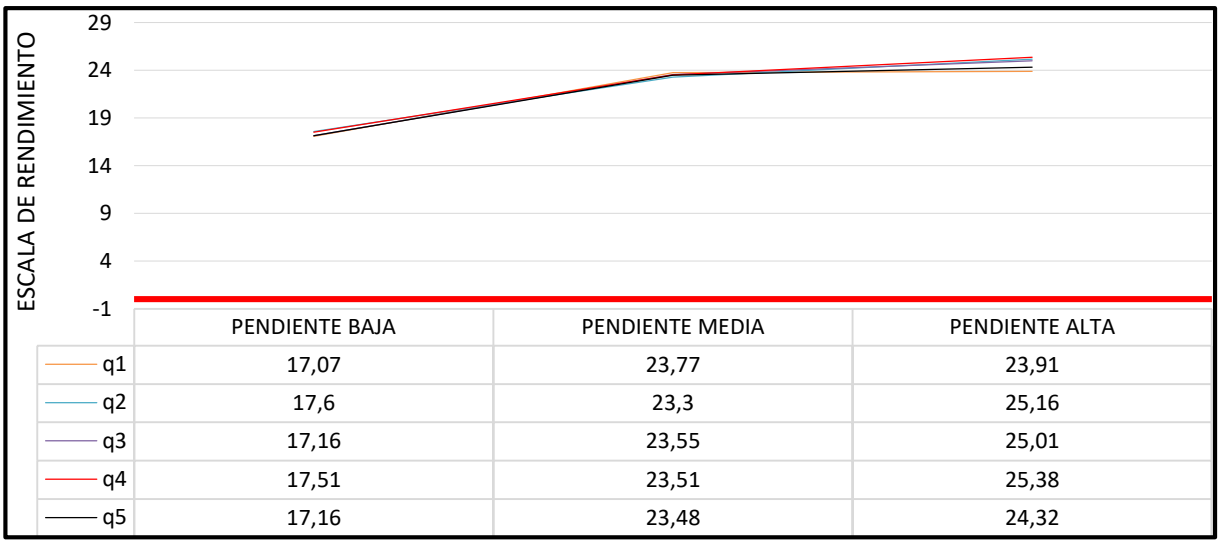
Cuando los estudiantes asisten a un aula con un alto residuo en la pendiente, como es el aula 218, también son los estudiantes del primer quintil los que menos incrementan su rendimiento (12,96, puntos), por el contrario, los situados en el quintil cinco, son los que más crecen, ya que, su rendimiento se incrementa 14 puntos.

Por lo tanto, a la vista de estos resultados, y tal y como se puede observar en el gráfico de la Tabla VI.65, los estudiantes del primer quintil son los que tienen menos ganancia cuando asisten a un aula situado en el quintil cuatro, independientemente del residuo de la pendiente del mismo. Sin embargo, los estudiantes agrupados en el quintil dos, son los que más incrementan su tasa de rendimiento, menos en el aula 218, donde el residuo de la pendiente es alto. En este caso, tal y como se aprecia en el gráfico, son los estudiantes del quintil cinco los que más crecen.

Finalmente, dentro de este modelo 14.2, se muestran los resultados de los estudiantes que asisten a aulas situadas en el último quintil.

Tabla VI.66. *Resultados del modelo 14.2 quintil 5*

MODELO 14.2 QUINTIL 5													
AULA	$Z(r_{0ij})\_q$	N	Y_{0ij}	Y_{8ij}	$Y_{8ij} - Y_{0ij}$	$Z(u_{00j})$	$Z(u_{10j})$	$Z(r_{0ij})$	u_{00j}	u_{10j}	r_{0ij}	e_{0ij}	e_{8ij}
88	1	5	7,23	24,3	17,07	1,39	**-1,72 (Baja)**	-1,54	3,27	-2,86	-1,86	0,33	0,06
	2	5	7,71	25,31	17,6			-0,64			-0,78	-0,28	-0,02
	3	5	8,71	25,87	17,16			-0,01			-0,02	-0,04	-0,22
	4	6	8,96	26,47	17,51			0,46			0,54	-0,34	-0,18
	5	2	10,25	27,41	17,16			0,87			1,04	0,44	0,27
292	1	5	5,52	29,29	23,77	1,48	**0,36 (Media)**	-1,46	3,48	0,63	-1,78	-0,19	0,37
	2	3	7,07	30,37	23,3			-0,41			-0,5	0,09	0,17
	3	5	7,32	30,87	23,55			0,15			0,17	-0,32	0,01
	4	8	8,16	31,67	23,51			0,54			0,64	0,04	0,33
	5	5	9,05	32,53	23,48			1,41			1,69	-0,13	0,14
218	1	3	5,94	29,85	23,91	1,53	**0,79 (Alta)**	-1,07	3,58	1,35	-1,3	-0,02	-0,86
	2	8	6,27	31,43	25,16			-0,61			-0,75	-0,25	0,17
	3	3	7,18	32,19	25,01			-0,01			-0,02	-0,06	0,2
	4	5	7,7	33,08	25,38			0,5			0,6	-0,17	0,47
	5	7	9,22	33,54	24,32			1,22			1,46	0,5	0,07



Nuevamente, todos los estudiantes incrementan su tasa de crecimiento, además, dentro de cada una de las aulas, los mismos tienen una ganancia en logro muy semejante, por ello en el gráfico todas las líneas se encuentran muy juntas. En las aulas del quintil cinco, del modelo 14, los estudiantes que asistían a un aula con una pendiente baja, no tenían un crecimiento positivo y dicha tasa entre los estudiantes situados en los diferentes quintiles era más variada.

En el aula 129, la tasa de crecimiento de los estudiantes de los distintos quintiles se ve incrementada alrededor de unos 17 puntos, o 17,5 puntos, siendo los estudiantes situados en el primer quintil los que muestran una menor tasa de

crecimiento (17,07 puntos,) y los del quintil cuatro, los que muestran una mayor ganancia (17,51 puntos).

Cuando los estudiantes asisten al aula 73, todos ellos incrementan su rendimiento promedio más de 23 puntos. En esta aula son los estudiantes del primer quintil los que más crecen, y los del segundo quintil los que menos.

Finalmente, los estudiantes que asisten a un aula con un residuo en la pendiente de crecimiento alto, como es el aula 333, incrementan su rendimiento más de 23 puntos. Ésta es la mejor aula de todas las aulas del modelo 14.2, ya que, es cuando más se incrementa la tasa de crecimiento de todos los estudiantes, independientemente del quintil al que pertenezcan.

En general, en las aulas de este quintil cinco (aulas 129, 73 y 333) son donde los estudiantes más crecen, en comparación con las aulas del resto de quintiles.

Entre llevar a un estudiante a un aula donde la fuerza de relación entre el intercepto y la pendiente es baja y positiva (modelo 14), y un aula donde la fuerza de relación entre dichos coeficientes es tambien baja, pero de direccionalidad negativa (modelo 14.2), la decisión más acertada, depende del quintil en el que se encuentre situada el aula.

Si se comparan los resultados de las aulas que se encuentran en el primer quintil del modelo 14 (la correlación entre el intercepto y la pendiente del nivel de aula se fijó a 0,2) con los del modelo 14.2 (la correlación entre el intercepto y la pendiente del nivel de aula se fijó a - 0,2), los estudiantes que sufren una menor pérdida entre la primera medida y la última, son los que asisten a aulas donde los que parten de un rendimiento alto, crecen mucho a lo largo del tiempo y, por el contrario, las que parten de un rendimiento bajo, su tasa de logro no se incrementa mucho a lo largo de las aplicaciones (modelo 14).

Si las aulas se encuentran en el quintil dos, depende del residuo de la pendiente de dicha aula, si el residuo es bajo, es mejor asistir a aulas que parten de un rendimiento alto y crecen poco o se mantienen constantes (modelo 14.2). Si el residuo del aula es medio o alto, es mejor seleccionar aulas con una correlación entre el intercepto y la pendiente positiva baja (modelo 14).

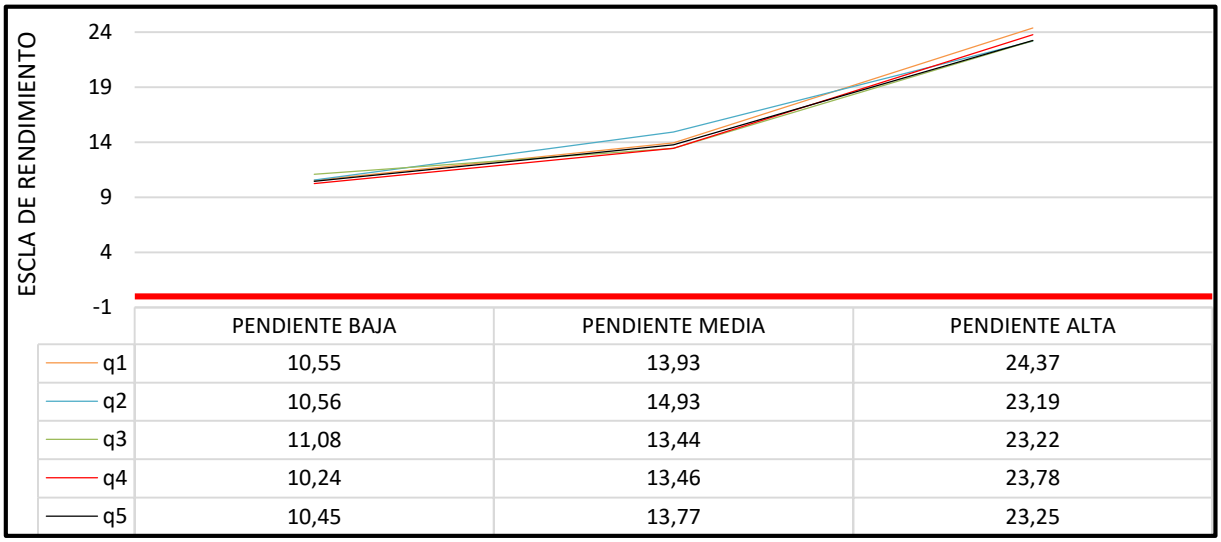
En las aulas del quintil tres, si el residuo de la pendiente del aula es bajo, es mejor seleccionar aulas donde la correlación entre el estatus y la pendiente del crecimiento de las aulas es baja de direccionalidad negativa, pero entre las aulas con un residuo medio o alto, los estudiantes incrementan más su tasa de crecimiento en las aulas del modelo 14.

Finalmente, las aulas situadas en el quintil cuatro y cinco, muestran la misma tendencia. Si los estudiantes asisten a aulas con un residuo en la pendiente bajo o medio, tienen una mayor ganancia en las aulas del modelo 14.2, pero, en las aulas con un residuo alto en la pendiente, su tasa de crecimiento entre la primera ocasión de medida y la última es ligeramente superior en las aulas del modelo 14.

A continuación, se muestran los resultados de los cinco quintiles referidos al modelo 14.3, en el cual se fijó una correlación entre el intercepto y la pendiente del nivel de aula de -0,9, por lo tanto, en dicho modelo existía una tendencia muy marcada en la cual, las aulas que parten de niveles de rendimiento alto se mantienen constantes, o no incrementan mucho su tasa de crecimiento a lo largo del tiempo, y por el contrario, si son aulas que parten de niveles de logro bajo, crecen mucho en las diferentes aplicaciones. En este modelo, también fue posible seleccionar las mismas aulas, para cada uno de los quintiles, que en el modelo 14.2.

Tabla VI.67. *Resultados del modelo 14.3 quintil 1*

MODELO 14.3 QUINTIL 1													
AULA	$Z(r_{0ij})\_q$	**N**	Y_{0ij}	Y_{8ij}	$Y_{8ij} - Y_{0ij}$	$Z(u_{00j})$	$Z(u_{10j})$	$Z(r_{0ij})$	u_{00j}	u_{10j}	r_{0ij}	e_{0ij}	e_{8ij}
240	1	5	-0,28	10,27	10,55	-1,46	**0,65 (Baja)**	-1,41	-3,19	0,93	-1,71	0,02	-0,08
	2	4	0,68	11,24	10,56			-0,66			-0,8	0,07	-0,01
	3	4	1,41	12,49	11,08			0,09			0,1	-0,11	0,34
	4	5	2,05	12,29	10,24			0,48			0,57	0,07	-0,33
	5	8	2,94	13,39	10,45			1,36			1,64	-0,11	-0,3
67	1	3	-0,55	13,38	13,93	-1,45	**0,95 (Media)**	-1,97	-3,17	1,4	-2,39	0,41	-0,06
	2	7	0,36	15,29	14,93			-0,64			-0,78	-0,29	0,25
	3	5	1,75	15,19	13,44			-0,04			-0,06	0,38	-0,58
	4	5	2,72	16,18	13,46			0,57			0,68	0,62	-0,33
	5	6	3,25	17,02	13,77			1,44			1,73	0,1	-0,54
188	1	2	-2,17	22,2	24,37	-1,43	**1,66 (Alta)**	-2,28	-3,13	2,47	-2,76	-0,88	0,5
	2	8	0,58	23,77	23,19			-0,42			-0,52	-0,37	-0,17
	3	8	1,23	24,45	23,22			0,04			0,04	-0,28	-0,05
	4	4	1,63	25,41	23,78			0,62			0,74	-0,59	0,21
	5	3	3,01	26,26	23,25			1,37			1,64	-0,1	0,16



En la tabla anterior VI.67, lo primero que se observa es que todos los estudiantes, independientemente del quintil y del tipo de aula al que asistan, obtienen una ganancia positiva en su puntuación de rendimiento, entre la primera ocasión de medida a la última. En el modelo 14.1, en el cual se fijó también una correlación de 0,9 pero con direccionalidad positiva, ocurría lo contrario, es decir, en todas las aulas seleccionadas, los estudiantes de los distintos quintiles, sufrían un decrecimiento de su rendimiento, siendo más pronunciado cuanto más pequeño era el residuo de la pendiente del aula al que asistían.

Si se comparan los valores de esta primera tabla del modelo 14.3, con los del quintil uno del modelo 14.2, también se observa unos resultados muy diferentes, y es que, en este último modelo todos los estudiantes, de los distintos quintiles y aulas,

obtenían una ganancia negativa entre la primera ocasión de medida y la última. Sin embargo, en el modelo 14.3, todos los estudiantes incrementan su tasa de rendimiento, incluso cuando asisten a un aula con un residuo bajo en la pendiente (0,65). En esta aula (240) los estudiantes crecen más de 10 puntos, siendo los situados en el cuarto quintil, los que menos incrementan su rendimiento (10,24) y los del quintil tres, los que más (11,08).

En el caso de los estudiantes que asisten al aula 67, donde el residuo de la pendiente es medio, todos los estudiantes de los distintos quintiles incrementan su rendimiento más de 13 puntos, siendo los del quintil dos los que más crecen (14,93 puntos).

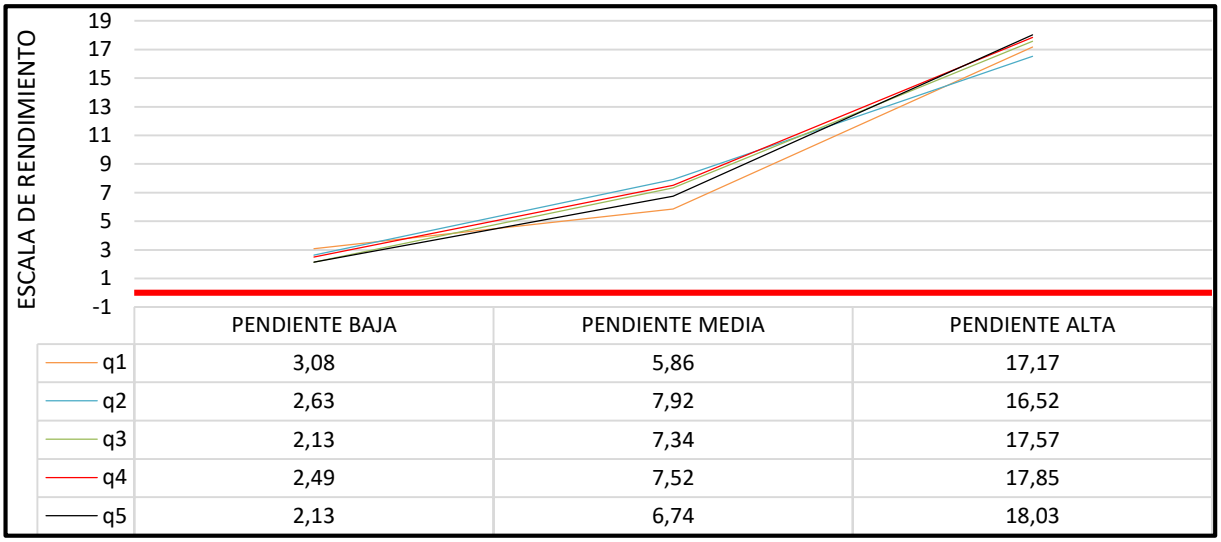
Finalmente, los estudiantes que asisten a un aula con un residuo alto en la pendiente (aula 188), su incremento es de más de 23 puntos en la escala de rendimiento.

Las diferencias encontradas entre el modelo 14.2 y el modelo 14.3 del quintil, se debe al efecto de la relación entre el estatus y la pendiente del nivel de aula. En el modelo 14.3, como se fijó una correlación negativa, pero de intensidad alta (-0,9), aunque los estudiantes se encuentren en aulas que, por su rendimiento promedio de partida se sitúan en el primer quintil, su ganancia es positiva, ya que, el valor de correlación fijado hace que predomine una tendencia en la cual, si las aulas parten de un rendimiento bajo su crecimiento se incrementa mucho a lo largo del tiempo. Esta tendencia no es tan marcada en el modelo 14.2, porque, aunque también se fija una correlación negativa, que conlleva que las aulas que parten de rendimiento bajo crezcan mucho, y las de rendimiento alto se mantengan más constantes, como el valor de la correlación tiene una intensidad baja (0,2), no es una tendencia que en promedio sea muy marcada.

La siguiente Tabla VI.68, muestra los resultados de las aulas seleccionadas dentro del segundo quintil.

Tabla VI.68. *Resultados del modelo 14.3 quintil 2*

MODELO 14.3 QUINTIL 2													
AULA	$Z(r_{0ij})$ _q	N	Y_{0ij}	Y_{8ij}	$Y_{8ij} - Y_{0ij}$	$Z(u_{00j})$	$Z(u_{10j})$	$Z(r_{0ij})$	u_{00j}	u_{10j}	r_{0ij}	e_{0ij}	e_{8ij}
139	1	9	1,76	4,84	3,08	-0,52	**-0,01 (Baja)**	-1,31	-1,05	-0,07	-1,59	-0,2	0,22
	2	3	2,77	5,4	2,63			-0,42			-0,52	-0,26	-0,29
	3	5	3,92	6,05	2,13			0			-0,01	0,39	-0,14
	4	8	3,92	6,41	2,49			0,52			0,62	-0,26	-0,42
	5	3	6,19	8,32	2,13			1,47			1,77	0,87	0,34
198	1	7	2,74	8,6	5,86	-0,48	**0,35 (Media)**	-1,1	-0,97	0,48	-1,33	0,45	-0,78
	2	6	2,99	10,91	7,92			-0,41			-0,5	-0,14	0,7
	3	5	3,63	10,97	7,34			0,01			0	0	0,26
	4	4	3,86	11,38	7,52			0,54			0,65	-0,41	0,02
	5	4	5,72	12,46	6,74			1,19			1,43	0,67	0,32
102	1	6	1,84	19,01	17,17	-0,48	**1,19 (Alta)**	-1,61	-0,97	1,76	-1,96	0,18	0,05
	2	4	2,85	19,37	16,52			-0,52			-0,63	-0,14	-0,92
	3	3	3,53	21,1	17,57			0,07			0,07	-0,17	0,11
	4	9	3,76	21,61	17,85			0,46			0,54	-0,41	0,15
	5	2	4,63	22,66	18,03			1,07			1,28	-0,27	0,47



Como ocurría en la Tabla VI.67, ahora en todas las aulas situadas en el segundo quintil todos los estudiantes incrementan su rendimiento promedio desde la primera ocasión de medida a la última, aunque dicho incremento es considerablemente menor en este quintil 2 del modelo 14.3. Esto se puede deber a que, en general, las aulas de este quintil, parten de un nivel de rendimiento más alto, que las del primer quintil y, debido al efecto de la correlación, su tasa de crecimiento entre la primera ocasión de medida y la última se incrementa menos.

Si se comparan los resultados de la tabla del quintil dos de este modelo 14.3, con los del modelo 14.1, la diferencia entre las mismas, no es tan marcada como en el caso anterior. Pues, aunque es verdad que los estudiantes del modelo 14.1, que asisten a un aula en el cual el residuo de su pendiente es bajo, sufren una disminución

en su tasa de crecimiento, los que asisten a un aula con un residuo en la pendiente medio o alto, obtenían una ganancia positiva como en el modelo 14.3, aunque es mayor en este último modelo.

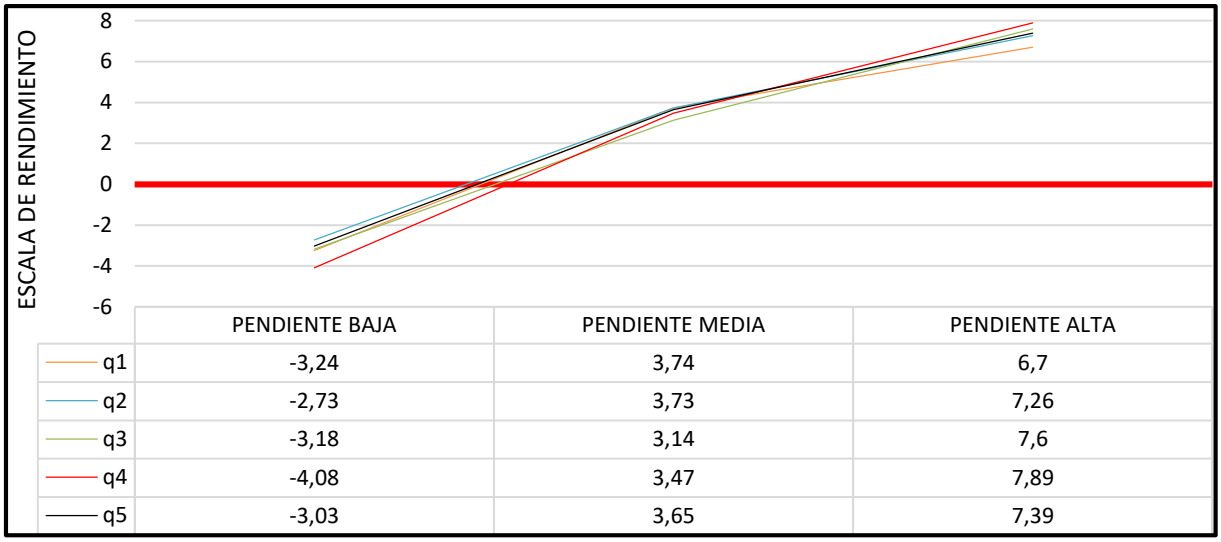
En las aulas del quintil dos del modelo 14.2, todos los estudiantes que asistían a un aula con una pendiente baja o media, obtenían una ganancia negativa, por lo tanto, en el modelo 14.3 los estudiantes obtienen mejores resultados. Lo mismo pasa cuando los estudiantes acuden a un aula con una pendiente alta, aunque en esta situación los estudiantes del modelo 14.2 tenían un crecimiento positivo, era muy inferior al del modelo 14.3.

En este modelo14.3, los estudiantes que asisten al aula 139 (residuo de la pendiente bajo, cercano a 0), incrementan su rendimiento entre 2 y 2,7 puntos, excepto los estudiantes del primer quintil que llegan a incrementarlo hasta 3 puntos. En el caso de los estudiantes que se encuentran en el aula 198, su crecimiento es superior a seis puntos, menos los estudiantes situados en el primer quintil, cuya tasa de crecimiento es de 5,86. Finalmente, los estudiantes que asisten a un aula donde el residuo de la pendiente es alto, como el aula 102, son los estudiantes del segundo quintil los que menos incrementan su rendimiento (16,52, puntos), y, los del quintil cinco los que más (18,03, puntos).

En cuanto a la tabla siguiente, se presentan los resultados de los estudiantes que se encuentran en aulas del tercer quintil.

Tabla VI.69. *Resultados del modelo 14.3 quintil 3*

MODELO 14.3 QUINTIL 3													
AULA	$Z(r_{0ij})$ _q	N	Y_{0ij}	Y_{8ij}	$Y_{8ij} - Y_{0ij}$	$Z(u_{00j})$	$Z(u_{10j})$	$Z(r_{0ij})$	u_{00j}	u_{10j}	r_{0ij}	e_{0ij}	e_{8ij}
55	1	7	3,55	0,31	-3,24	0,04	**-0,48 (Baja)**	-1,5	0,22	-0,8	-1,82	0,54	0,48
	2	5	3,64	0,91	-2,73			-0,57			-0,7	-0,48	-0,04
	3	6	4,63	1,45	-3,18			-0,08			-0,11	-0,08	-0,09
	4	3	6,41	2,33	-4,08			0,79			0,94	0,65	-0,26
	5	4	6,6	3,57	-3,03			1,72			2,07	-0,29	-0,15
297	1	5	2,4	6,14	3,74	0,05	**0,07 (Media)**	-1,94	0,23	0,05	-2,35	-0,07	0,07
	2	4	4,21	7,94	3,73			-0,44			-0,54	-0,08	0,05
	3	3	5,13	8,27	3,14			0,03			0,03	0,28	-0,18
	4	9	5,52	8,99	3,47			0,49			0,58	0,11	-0,02
	5	7	6,27	9,92	3,65			1,43			1,72	-0,28	-0,23
192	1	5	3,61	10,31	6,7	0,06	**0,4 (Alta)**	-1,41	0,25	0,55	-1,71	0,47	-0,47
	2	5	4,39	11,65	7,26			-0,5			-0,62	0,15	-0,22
	3	5	4,47	12,07	7,6			0,03			0,02	-0,4	-0,44
	4	5	5,45	13,34	7,89			0,63			0,75	-0,16	0,1
	5	4	6,76	14,15	7,39			1,52			1,83	0,09	-0,16



Por primera vez, dentro de los quintiles del modelo 14.3, tal y como se puede observar en la tabla anterior VI.69, los estudiantes que asisten a un aula con un residuo en la pendiente bajo, experimentan una ganancia negativa en la escala de rendimiento. Esto también ocurría con los estudiantes que asistían a un aula con un residuo bajo, del modelo 14.1, ya que, los estudiantes de todos los quintiles de dicha aula disminuían su rendimiento en más de 3 puntos, excepto los del quintil cuatro cuya tasa de rendimiento decrecía 1,61 puntos. En cuanto al rendimiento del resto de estudiantes de las distintas aulas, en el modelo 14.1, los estudiantes situados en un aula con un residuo en la pendiente medio y alto, obtenian una tasa de rendimiento superior al del modelo 14.3. También es superior la tasa de crecimiento de todos los estudiantes que asisten a aulas con un residuo en la pendiente bajo y medio del

modelo 14.2, respecto al modelo 14.3, pero si el residuo del aula es alto (aula 192) la tasa de crecimiento es ligeramente superior en el modelo 14.3.

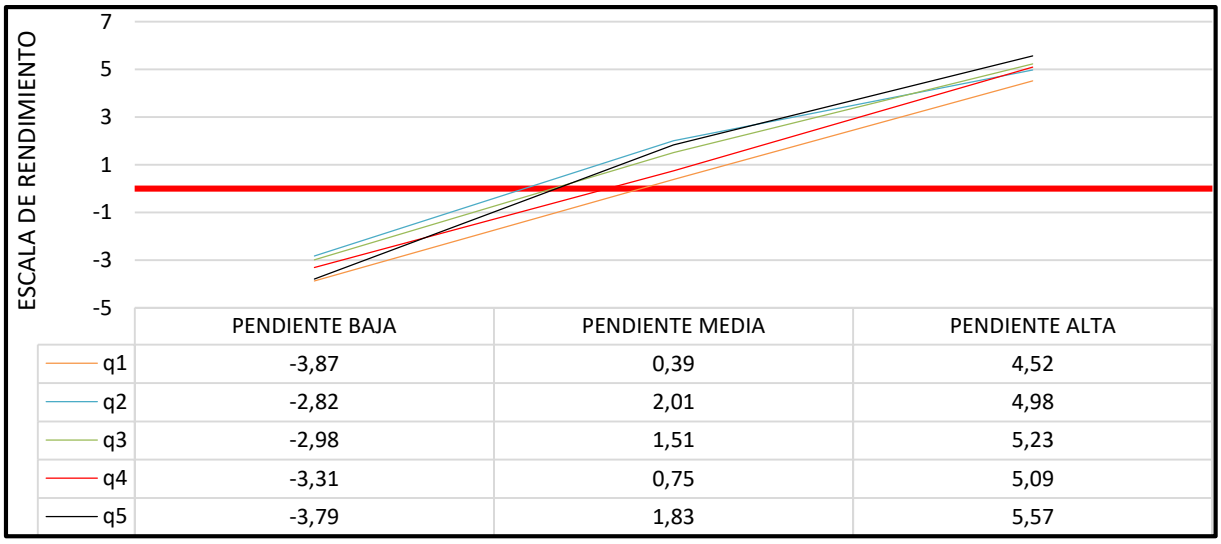
En la Tabla VI.69, se puede observar que los estudiantes que asisten al aula 55, su rendimiento decrece entre 2 y 4 puntos, siendo los estudiantes situados en el segundo quintil los que menos se ven afectados por dicha disminución, ya que, su rendimiento decrece 2,73 puntos y, sin embargo, los estudiantes del cuarto quintil los que más, su rendimiento decrece 4,08 puntos.

En cuanto a los estudiantes que asisten al aula 297, cuyo residuo en la pendiente es medio, incrementan su rendimiento alrededor de 3-4 puntos, y los del aula 192, su tasa de crecimiento entre la primera ocasión de medida y la última es superior a 7 puntos, a excepción de los estudiantes situados en el primer quintil que lo incrementan en 6,7 puntos.

En la siguiente tabla se presentan los resultados de los estudiantes que se encuentran situados en el cuarto quintil.

Tabla VI.70. *Resultados del modelo 14.3 quintil 4*

MODELO 14.3 QUINTIL 4													
AULA	$Z(r_{0ij})\_q$	N	Y_{0ij}	Y_{8ij}	$Y_{8ij} - Y_{0ij}$	$Z(u_{00j})$	$Z(u_{10j})$	$Z(r_{0ij})$	u_{00j}	u_{10j}	r_{0ij}	e_{0ij}	e_{8ij}
88	1	3	4,69	0,82	-3,87	0,51	-0,48 (Baja)	-1,58	1,28	-0,8	-1,92	0,73	0,03
	2	7	5,09	2,27	-2,82			-0,56			-0,69	-0,1	0,24
	3	5	5,81	2,83	-2,98			0,02			0,01	-0,08	0,1
	4	6	6,19	2,88	-3,31			0,51			0,6	-0,29	-0,43
	5	6	7,82	4,03	-3,79			1,39			1,67	0,27	-0,36
292	1	5	4,53	4,92	0,39	0,51	-0,11 (Media)	-1,4	1,29	-0,23	-1,71	0,35	-0,6
	2	5	4,92	6,93	2,01			-0,54			-0,66	-0,3	0,37
	3	3	6,11	7,62	1,51			0,08			0,09	0,14	0,31
	4	4	6,36	7,11	0,75			0,46			0,55	-0,07	-0,66
	5	6	7,16	8,99	1,83			1,39			1,67	-0,4	0,09
218	1	3	4,98	9,5	4,52	0,52	0,18 (Alta)	-1,1	1,3	0,22	-1,34	0,43	-0,04
	2	5	5,34	10,32	4,98			-0,56			-0,68	0,12	0,11
	3	4	6,05	11,28	5,23			0,07			0,08	0,07	0,32
	4	7	6,38	11,47	5,09			0,55			0,66	-0,18	-0,08
	5	6	7,72	13,29	5,57			1,76			2,11	-0,29	0,29



En el modelo 14.3 del quintil cuatro, ocurre lo mismo que en las aulas del quintil tres del modelo 14.3, y es que, los estudiantes que asisten a un aula con un residuo en la pendiente bajo, experimentan una ganancia negativa entre la primera ocasión de medida y la última. Mientras que los estudiantes que asisten a aulas con un residuo en la pendiente medio o alto, la ganancia de los mismos es positiva, aunque los estudiantes del aula 292, incrementan su tasa de crecimiento muy poco, sobre todo, los estudiantes que se sitúan en el primer quintil y en el cuarto, ya que, su tasa de rendimiento entre la primera ocasión de medida y la última no se incrementa más de un punto. Son los estudiantes situados en el segundo quintil los que más incrementan su tasa de rendimiento dentro de dicha aula (2 puntos). Finalmente, en

cuanto a los estudiantes que asisten a un aula con un residuo en la pendiente que es alto (aula 218), su rendimiento crece entre 4,52 y 5,57.

En las tres aulas de dicho quintil, son los estudiantes situados en el primer quintil los más perjudicados, ya que, son los que más pérdida tienen y los que menos ganancia reciben. Sin embargo, entre los más beneficiados, se encuentran los estudiantes situados en el quintil dos, ya que, son los que menos pérdida experimentan y los que más incrementan su tasa de rendimiento, sobre todo, cuando el residuo de la pendiente es medio.

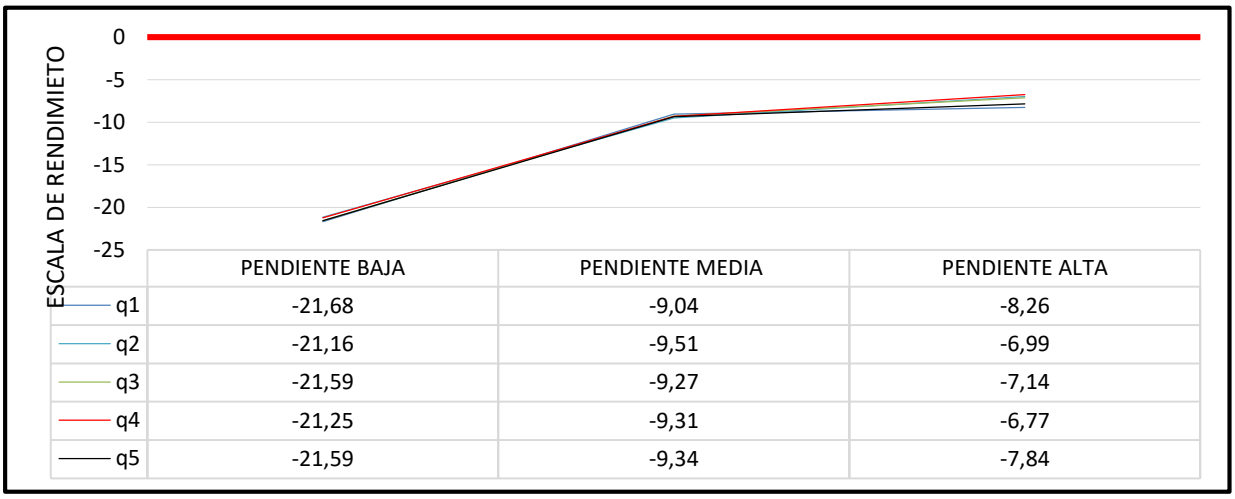
Si se comparan los resultados de la tabla anterior con los resultados del quintil cuatro del modelo 14.1, en el cual se fijó una correlación entre el intercepto y la pendiente de 0,9, se obtienen unos resultados muy diferentes, ya que, en este último modelo todos los estudiantes de las diferentes aulas seleccionadas tienen una ganancia positiva. Siendo su tasa de crecimiento superior a 10 puntos, cuando asisten a un aula con un residuo medio en la pendiente y superior a 14 puntos, cuando asisten a un aula con un residuo alto en la pendiente.

Al comparar los resultados de la Tabla VI.70 (modelo 14.3, quintil 4) con los de la Tabla VI.65, referida a las aulas del quintil cuatro del modelo 14.2 (la correlación entre el intercepto y la pendiente del nivel de aula se a fijó -0,2), se obtienen resultados muy diferentes. En este último modelo 14.2, todos los estudiantes incrementan su tasa de crecimiento desde la primera ocasión de medida a la última. Los estudiantes que asisten al aula 88, la incrementan más de 9 puntos, los del aula 292, más de 10 puntos y los del aula 218, más de 12 puntos.

Por último, se presentan los resultados de los estudiantes que asisten a aulas situadas en el quinto quintil.

Tabla VI.71. *Resultados del modelo 14.3 quintil 5*

MODELO 14.3 QUINTIL 5													
AULA	$Z(r_{0ij})\_q$	N	Y_{0ij}	Y_{8ij}	$Y_{8ij} - Y_{0ij}$	$Z(u_{00j})$	$Z(u_{10j})$	$Z(r_{0ij})$	u_{00j}	u_{10j}	r_{0ij}	e_{0ij}	e_{8ij}
129	1	5	6,38	-15,3	-21,68	1,41	**-1,97 (Baja)**	-1,54	3,32	-3,08	-1,86	0,33	0,06
	2	5	6,87	-14,29	-21,16			-0,64			-0,78	-0,28	-0,02
	3	5	7,86	-13,73	-21,59			-0,01			-0,02	-0,04	-0,22
	4	6	8,12	-13,13	-21,25			0,46			0,54	-0,34	-0,18
	5	2	9,4	-12,19	-21,59			0,87			1,04	0,44	0,27
73	1	5	6,1	-2,94	-9,04	1,47	**-1,01 (Media)**	-1,46	3,46	-1,6	-1,78	-0,19	0,37
	2	3	7,65	-1,86	-9,51			-0,41			-0,5	0,09	0,17
	3	5	7,91	-1,36	-9,27			0,15			0,17	-0,32	0,01
	4	8	8,74	-0,57	-9,31			0,54			0,64	0,04	0,33
	5	5	9,63	0,29	-9,34			1,41			1,69	-0,13	0,14
333	1	3	6,83	-1,43	-8,26	1,51	**-0,83 (Alta)**	-1,07	3,55	-1,33	-1,3	-0,02	-0,86
	2	8	7,15	0,16	-6,99			-0,61			-0,75	-0,25	0,17
	3	3	8,06	0,92	-7,14			-0,01			-0,02	-0,06	0,2
	4	5	8,58	1,81	-6,77			0,5			0,6	-0,17	0,47
	5	7	10,11	2,27	-7,84			1,22			1,46	0,5	0,07



En la Tabla VI.71, lo primero que se observa, es que, por primera vez dentro de este modelo 14.3, todos los estudiantes tienen una ganancia negativa entre la primera ocasión de medida y la última, independientemente del valor del residuo de la pendiente de crecimiento del aula al que asisten y, del quintil en el cual se encuentren agrupados. En este quintil cinco, ocurre lo contrario al primer quintil, donde todos los estudiantes incrementaban su tasa de crecimiento, por ello, la explicación para este quintil cinco, es la misma que la dada para las aulas del primer quintil, pero con la lógica contraria, es decir, en este modelo 14.3, los estudiantes situados en aulas del último quintil, sufren una disminución de su tasa de crecimiento desde la primera ocasión de medida a la última, debido a que, asisten a aulas donde en promedio el

rendimiento de partida es muy alto. Al fijarse entre el intercepto y la pendiente de dicho nivel, una correlación alta y negativa (-0,9), en general, las aulas que parten de un rendimiento promedio alto, tenderán a disminuir el mismo a lo largo del tiempo, o bien, su crecimiento se mantendrá bastante estable, por el efecto de la correlación.

Sin embargo, en las aulas del quintil cinco del modelo 14.2, no ocurría esto, sino que, todos los estudiantes mostraban una ganancia positiva, pues, aunque la correlación entre el intercepto y la pendiente del nivel de estudiante también era negativa, la intensidad de relación era baja (0,2). Del mismo modo, los estudiantes de las aulas del quintil cinco del modelo 14.1, donde la correlación se fijó a 0,9, también obtenían una ganancia positiva en su tasa de crecimiento, aunque dicha ganancia era inferior a la del modelo 14.2, a pesar de que el valor de la correlación en el modelo 14.1, se fijó a 0,9.

Acorde con los resultados del modelo 14.3, del quintil cinco, en esta situación, son los estudiantes del primer quintil los que, en general, menos beneficiados se encuentran al asistir a aulas formadas por estudiantes que en su mayoría cuentan con un rendimiento de partida alto. Sin embargo, en el modelo 14.1, donde la correlación entre el intercepto y la pendiente era también alta, pero con direccionalidad positiva, ocurría lo contrario, ya que, estos estudiantes experimentaban una ganancia importante en su tasa de crecimiento cuando asistian a aulas con un rendimiento promedio inicial alto.

Si se tuviera que elegir entre aulas compensatorias donde existe una relación negativa entre el intercepto y la pendiente, si las aulas pertenecen al quintil uno o dos, la tasa de crecimiento de los estudiantes a lo largo del tiempo será más elevada cuando asisten a aulas donde el rendimiento de partida no se encuentra muy relacionado con el adquirido a lo largo del tiempo (modelo 14.3). Sin embargo, si las aulas se situan en el quintil tres, los estudiantes cuando asisten a aulas con una pendiente baja o media crecen más en las aulas compensatorias donde dicha tendencia no es muy marcada (correlación entre el intercepto y la pendiente de -0,2) pero si el aula cuenta con un residuo en la pendiente alto parece no existir tanta diferencia entre asistir a aulas donde la relación entre el estatus y la pendiente es de intensidad negativa alta (modelo 14.3) o baja (modelo 14.2). Finalmente, en cuanto a las aulas situadas en el quintil cuatro y cinco los estudiantes incrementan más sus niveles de logro entre la primera ocasión de medida y el resto, en el modelo 14.2.

VI.2. Problema 2. Influencia del tamaño muestral de las aulas en la estimación de su pendiente de crecimiento

El segundo problema de investigación se centraba en el estudio del residuo de crecimiento de las aulas, con el fin de conocer la influencia del tamaño de las mismas en la precisión con la que se estima su pendiente de crecimiento. En definitiva, se trataba de comprobar, si las aulas compuestas por pocos estudiantes como son menos de 10 (aulas pequeñas), o menos de 20 (aulas medianas), podían diferenciarse de la media de crecimiento, o, por el contrario, su pendiente de crecimiento se estima con tanto error que se requiere de un mayor número de estudiantes por aula, más de 19 (aulas grandes,) para poder conocer su crecimiento.

Dentro de este segundo problema de investigación, recuérdese que también se plantearon dos objetivos específicos.

Mediante el primer objetivo específico, se pretendía garantizar una adecuada recuperación de las condiciones experimentales fijadas, antes de la generación definitiva de los resultados, a través de la simulación de distintos porcentajes de tamaños de aula (pequeños, medianos y grandes), manteniendo los valores de correlación y de varianza fijados en el modelo 15 del primer problema de investigación.

En el segundo objetivo específico se plantea estudiar, en tres situaciones distintas (*situaciones a, b* y *c*), donde se simulan diferentes porcentajes de aulas pequeñas (menos de 11 estudiantes), medianas (menos de 20 estudiantes) y grandes (más de 20 estudiantes), cómo afecta el tamaño de las aulas, en estas tres situaciones, en la precisión con la que se estima el crecimiento de las mismas.

Los resultados obtenidos en cada uno de estos objetivos específicos son los que se presentan a continuación.

VI.2.1. Objetivo específico 1. Garantizar una decuada recuperación de las condiciones experimentales manipuladas

En este apartado solo se muestra la precisión con la que se recuperan las condiciones experimentales relacionadas con la generación de la muestra, ya que son el foco de interés en este problema de investigación. En cuanto a la recuperación de los coeficientes aleatorios y de los valores de correlación fijados entre el intercepto y

la pendiente, muestran el mismo comportamiento que el explicado en el primer objetivo del primer problema de investigación (*ver apartado VI.1.1.2*.), pero si se desea, se puede consultar la recuperación de estos parámetros, en cada uno de los tres modelos simulados, (modelo 15.a, modelo 15.b y modelo 15.c) dentro de la *situación a, b* y *c*, en el Anexo II, teniendo en cuenta que los valores de estos coeficientes son los mismos que los fijados para el modelo 15 del primer problema de investigación, y lo único que cambia es el número de aulas que se generan y el tamaño de las mismas.

En las siguientes tablas (Tabla VI.72, Tabla VI.73 y Tabla VI.74), se muestran los resultados obtenidos en la recuperación de los tamaños muestrales dentro de las distintas situaciones.

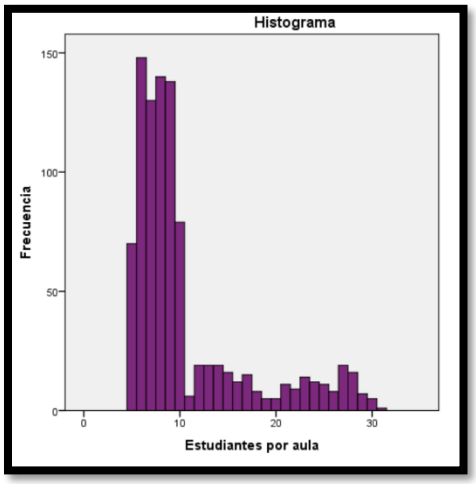
VI.2.1.1. Recuperación de los parámetros muestrales generados en los modelos de la situación a

En la ***situación a*** o **modelo 15.a**, se fijó la condición de que alrededor del 75% de la muestra total de aulas fueran pequeñas, es decir, que estuvieran comprendidas por un mínimo de 5 estudiantes y un máximo de 10. Además, se quería que el 25% de las aulas restantes estuvieran formadas en torno a un 12,5% por aulas medianas (compuestas por un mínimo de 11 estudiantes y un máximo de 19), y el otro 12,5% por aulas grandes (formadas por un minímo de 20 estudiantes y un máximo de 30), siendo la muestra total de estudiantes 10.000.

En la tabla siguiente VI.72 se muestra el número de estudiantes que hay dentro de cada aula, el número de aulas generadas con un determinado número de estudiantes y el resultado de multiplicar el número de estudiantes por el número de aulas, todo ello para cada tipo de aula (pequeñas, medianas y grandes). Además, esta tabla al igual que la del primer problema de investigación (Tabla VI.1), se acompaña de un histograma que representa la distribución muestral de la *situación a* o modelo 15.a

Tabla VI.72. *Recuperación de los tamaños muestrales de la situación a*

AULAS PEQUEÑAS			AULAS MEDIANAS			AULAS GRANDES		
Número estudiantes por aula	**Número aulas**	**Número estudiantes *N aulas**	**Número estudiantes por aula**	**Número aulas**	**Número estudiantes *N aulas**	**Número estudiantes por aula**	**Número aulas**	**Número estudiantes *N aulas**
5	70	350	**11**	6	66	**20**	5	100
6	148	888	**12**	19	228	**21**	11	231
7	130	910	**13**	19	247	**22**	9	198
8	140	1120	**14**	19	266	**23**	14	322
9	138	1242	**15**	16	240	**24**	12	288
10	79	790	**16**	12	192	**25**	11	275
			17	15	255	**26**	8	208
			18	8	144	**27**	19	513
			19	5	95	**28**	16	448
						29	7	203
						30	5	150
						31	1	31
N Total	**705 (74,8%)**	**5300 (53%)**	**N Total**	**119 (12,7%)**	**1733 (17,3%)**	**N Total**	**118 (12,5%)**	**2967 (29,7%)**



En la tabla anterior VI.72, se observa que se han generado un 74,8% de aulas pequeñas, un 12,7% de aulas medianas y un 12,5% de aulas grandes, siendo el total de aulas generadas 942, comprendidas por un mínimo de 5 estudiantes y un máximo de 31, donde la suma total de estudiantes es 10.000. A la vista de estos resultados se puede concluir que existe una buena recuperación de las condiciones experimentales establecidas, ya que los porcentajes de aulas simulados son muy cercanos a los fijados. También el número de estudiantes mínimo y máximo por aula, es básicamente el establecido. Tan solo en las aulas grandes se genera un aula comprendida por 31 estudiantes, cuando el máximo fijado fue de 30, pero esto puede ser debido a que la distribución se genera al azar. En cualquier caso, lo importante de las aulas grandes no es tanto el número máximo de estudiantes generados, que, por supuesto debe fijarse acorde a la legislación educativa, según la cual en secundaria

debe haber un máximo de 30 estudiantes por aula con posibilidad de ampliarse hasta 33 (LOMCE, 2016), sino que el mínimo de estudiantes en este tipo de aulas sea de 20.

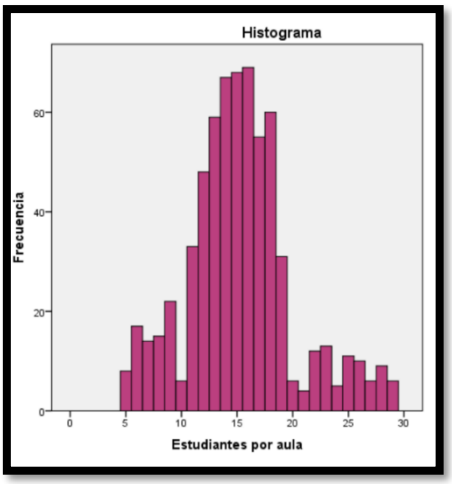
VI.2.1.2. Recuperación de los parámetros muestrales generados en los modelos de la situación b

En la ***situación b* o modelo 15.b,** se fijó como condición que alrededor del 75% de las aulas, fueran aulas medianas y el resto de aulas fueran pequeñas (12,5%), y grandes (12,5%), con un total de 10.000 estudiantes.

A continuación, se muestran los resultados muestrales que se generaron dentro de esta situación.

Tabla VI.73. *Recuperación de los tamaños muestrales de la situación b*

AULAS PEQUEÑAS			AULAS MEDIANAS			AULAS GRANDES		
Número estudiantes por aula	**Número aulas**	**Número estudiantes *N aulas**	**Número estudiantes por aula**	**Número aulas**	**Número estudiantes *N aulas**	**Número estudiantes por aula**	**Número aulas**	**Número estudiantes *N aulas**
5	8	40	**10**	6	60	**20**	6	120
6	17	102	**11**	33	363	**21**	4	84
7	14	98	**12**	48	576	**22**	12	264
8	15	120	**13**	59	767	**23**	13	299
9	22	198	**14**	67	938	**24**	5	120
			15	68	1020	**25**	11	275
			16	69	1104	**26**	10	260
			17	55	935	**27**	6	162
			18	60	1080	**28**	9	252
			19	31	589	**29**	6	174
N Total	**76 (11,6%)**	**558 (5,58%)**	**N Total**	**496 (75,8%)**	**7432 (74,3%)**	**N Total**	**82 (12,5%)**	**2010 (20,1%)**



Los resultados de la *situación b*, revelan que las condiciones muestrales fijadas se generan de manera precisa, ya que se han simulado un total de 10.000 estudiantes repartidos en 654 aulas, de las cuales, un 11,6% son pequeñas (menos de 11 estudiantes), un 75,8% son medianas (menos de 20 estudiantes), y un 12,5% son

grandes (más de 20 estudiantes). En estas aulas grandes se han generado un máximo de 29 estudiantes, sin llegar a los 30 fijados para este tipo de aula, esto puede deberse, nuevamente, al azar, pero no es un factor que afecte a los resultados finales.

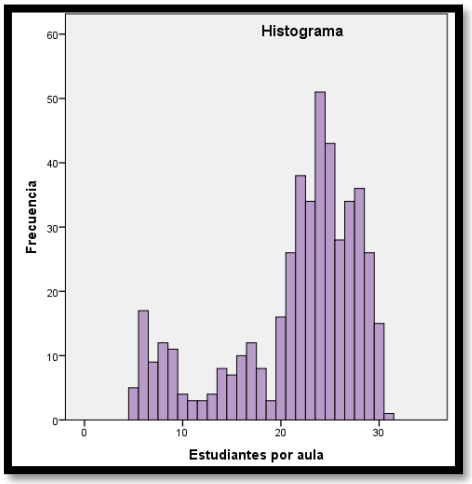
VI.2.1.1. Recuperación de los parámetros muestrales generados en los modelos de la situación c

Finalmente, en la ***situación c*** o **modelo 15.c**, se pretendía generar un total de 10.000 estudiantes repartidos en un 12,5% en aulas pequeñas, un 12,5% en aulas medianas y 75% en aulas grandes.

Los resultados obtenidos en la distribución muestral de esta situación, se muestran a continuación.

Tabla VI.74. *Recuperación de los tamaños muestrales de la situación c*

AULAS PEQUEÑAS			AULAS MEDIANAS			AULAS GRANDES		
Número estudiantes por aula	**Número aulas**	**Número estudiantes *N aulas**	**Número estudiantes por aula**	**Número aulas**	**Número estudiantes *N aulas**	**Número estudiantes por aula**	**Número aulas**	**Número estudiantes *N aulas**
5	5	25	**10**	4	40	20	16	320
6	17	102	**11**	3	33	21	26	546
7	9	63	**12**	3	36	22	38	836
8	12	96	**13**	4	52	23	34	782
9	11	99	**14**	8	112	24	51	1224
			15	7	105	25	43	1075
			16	10	160	26	28	728
			17	12	204	27	34	918
			18	8	144	28	36	1008
			19	3	57	29	26	754
						30	15	450
						31	1	31
N Total	**54 (11,6%)**	**385 (4%)**	**N Total**	**62 (13,4%)**	**943 (8,5%)**	**N Total**	**348 (75%)**	**8672 (87,5%)**



En la *situación c*, tal y como se muestra en la tabla anterior VI.74, los 10.000 estudiantes generados se encuentran repartidos en un total de 464 aulas, de las cuales el 75% son aulas grandes, formadas por un máximo de 31 estudiantes y un mínimo de 20, el 13,4% son aulas medianas y el 11,6% restantes son aulas pequeñas, por lo que, por lo que en general, las condiciones fijadas previamente se generan correctamente.

A la vista de todos estos resultados se puede concluir que, las condiciones experimentales fijadas para la simulación de la muestra en cada una de las diferentes situaciones se generan de manera adecuada, pues, aunque los porcentajes obtenidos en cada tipo de aula no son exactamente los fijados, debido al azar, se encuentran muy próximos a los previamente establecidos. En cuanto al tamaño muestral, tan sólo en las aulas grandes no se simula el tamaño máximo establecido, ya que se fijó un máximo de 30 estudiantes por aula, y en la *situación a* y *c* se generaron 31 estudiantes y en la *situación b*, 29. Tal y como se ha comentado, esto también puede ser debido al azar, pero en cualquier caso no afecta significativamente en el estudio de los tamaños muestrales.

VI.3.1. Objetivo específico 2. Estudio del efecto del tamaño muestral de las aulas en su pendiente decrecimiento

Con el fin de conocer que aulas (pequeñas, medianas o grandes), en cada una de las tres situaciones definidas previamente (*situación a, b y c*), su crecimiento se distingue de la media general, se realizó una tabla de contingencia a partir del residuo de la pendiente de crecimiento de las mismas, eliminando aquellas aulas que su residuo se encontraba en el 10% de su desviación típica (entre -0,07 y 0,07), por considerarse que se podían situar en la media no por el error de estimación sino por su crecimiento real.

En la primera Tabla.VI.75 se muestra el número de aulas iniciales existentes en cada situación, y el número final tras eliminar aquellas aulas que su residuo se encontraba entre el 10% de su desviación típica.

En la Tabla.VI.76 se presenta el porcentaje final de cada tipo de aula dentro de la *situación a, b* y *c,* comparado con el porcentaje incial.

Tabla VI.75. *Resumen de los casos*

	Casos válidos inciales		Caos válidos tras la restricción	
	N	Porcentaje	N	Porcentaje
a	942	100%	882	93,6%
b	654	100%	611	93,4%
c	464	100%	425	91,6%%

En la Tabla VI.75. *Resumen de los casos,* se observa que tras eliminar aquellas aulas cuyo residuo en la pendiente de crecimiento se encontraba comprendido entre -0,07 y 0,07, en la *situación a*, se trabajó con un total de 882 aulas, por lo que se eliminarón 60 aulas. En el caso de *la situación b*, inicialmente el total de aulas era 654 y tras la restricción realizada se redujeron a 611, por lo tanto, se eliminaron 43 aulas, y, finalmente, en *la situación c* se quitaron 39 aulas.

Tabla VI.76. *Porcentaje de aulas iniciales y finales en cada situación*

SITUACIÓN		Valores Iniciales		Valores tras la restricción	
		Frecuencia	Porcentaje	Frecuencia	Porcentaje
a	Aulas pequeñas	705	**74,8%**	657	**74,5%**
	Aulas medianas	119	**12,7%**	108	**12,2%**
	Aulas grandes	118	**12,5%**	117	**13,3%**
	Total	**942**	100%	**882**	100%
b	Aulas pequeñas	76	**11,6%**	78	**12,8%**
	Aulas medianas	496	**75,8%**	457	**74,8%**
	Aulas grandes	82	**12,5%**	76	**12,4%**
	Total	**654**	100%	**611**	100%
c	Aulas pequeñas	54	**11,6%**	54	**12,7%**
	Aulas medianas	62	**12,7%**	48	**11,3%**
	Aulas grandes	348	**75%**	323	**76,0%**
	Total	**464**	100%	**425**	100%

En la Tabla VI.76. *Porcentajes de aulas iniciales y finales en cada situación,* como su propio título indica se especifica el porcentaje inicial y final de cada tipo de aula en cada situación. En cuanto a los porcentajes finales, en la ***situación a,*** ahora el 74,5 % de las aulas son pequeñas, el 12,2% medianas y el 13,3% grandes. En la ***situación b***, un 12,8% son pequeñas, un 74,8% medianas y 12,4% grandes. En cuanto a la

situación c, el 76% de las aulas son grandes, el 11,3% medianas y el 12,7% pequeñas. Si se comparan estos porcentajes con los iniciales, se observa que la restricción de eliminar aquellas aulas que su residuo se encontraba entre el 10% de su desviación típica no afecta significativamente a los porcentajes finales.

En la siguiente tabla (VI.77), apartir de la muestra final de aulas, se presentan los resultados obtenidos en la tabla de contingencia con el fin de conocer dentro de cada situación, el porcentaje de aulas que se diferencian de la media acorde al tipo de aula.

Tabla VI.77. *Porcentaje de aulas que se diferencian de la media*

Situación				**0=No se diferencian de la media**	**1= Se diferencian de la media**	**Total**
a	Tamaño medio del aula	1 Aulas pequeñas	Recuentro	567	90	657
			% dentro del tamaño medio del aulas	**86,3%**	**13,7%**	100,0%
	Tamaño medio del aula	2 Aulas medianas	Recuentro	71	37	108
			% dentro del tamaño medio del aulas	**65,7%**	**34,3%**	100,0%
	Tamaño medio del aula	3 Aulas grandes	Recuentro	61	56	117
			% dentro del tamaño medio del aulas	**52,1%**	**47,9%**	100,0%
	Total		Recuentro	699	183	882
			% dentro del tamaño medio del aulas	**79,3%**	**20,7%**	100,0%
b	Tamaño medio del aula	1 Aulas pequeñas	Recuentro	66	12	78
			% dentro del tamaño medio del aulas	**84,6%**	**15,4%**	100,0%
	Tamaño medio del aula	2 Aulas medianas	Recuentro	320	137	457
			% dentro del tamaño medio del aulas	**70,0%**	**30,0%**	100,0%
	Tamaño medio del aula	3 Aulas grandes	Recuentro	42	34	76
			% dentro del tamaño medio del aulas	**55,3%**	**44,7%**	100,0%
	Total		Recuentro	428	183	611
			% dentro del tamaño medio del aulas	**70,0%**	**30,0%**	100,0%
c	Tamaño medio del aula	1 Aulas pequeñas	Recuentro	48	6	54
			% dentro del tamaño medio del aulas	**88,9%**	**11,1%**	100,0%
	Tamaño medio del aula	2 Aulas medianas	Recuentro	39	9	48
			% dentro del tamaño medio del aulas	**81,3%**	**18,8%**	100,0%
	Tamaño medio del aula	3 Aulas grandes	Recuentro	191	132	323
			% dentro del tamaño medio del aulas	**59,1%**	**40,9%**	100,0%
	Total		Recuentro	278	147	425
			% dentro del tamaño medio del aulas	**65,4%**	**34,6%**	100,0%

En la tabla anterior VI.77, dentro de la ***situación a***, donde recuérdese que el 74,5% de las aulas eran pequeñas (menos de 11 estudiantes por aula), el 12,2% medianas (menos de 20 estudiantes por aula) y un 13,3% grandes (más de 19 estudiantes por aula), se observa que en las aulas pequeñas el 86,3% no se diferencian de la media, en las aulas medianas este porcentaje se ve reducido a un

66% y en las aulas grandes a un 52,1%. En cuanto al número total de aulas que se diferencian de la media en esta situación es del 20,7%.

En la ***situación b***, donde el 74,8% de las aulas eran medianas, y el porcentaje restante eran aulas pequeñas (12,8%) y grandes (12,4%), siguen siendo las aulas pequeñas donde existe un mayor número que no se diferencian de la media (alrededor del 85%). Por el contrario, son las compuestas por más de 19 estudiantes, donde existe un mayor porcentaje de aulas (44,7%) en las cuales es posible conocer su pendiente de crecimiento, ya que se diferencian de la media. Además, respecto a la situación anterior se ha producido un incremento importante (de 10 puntos porcentuales) del número total de aulas en las cuales es posible conocer su pendiente media de crecimiento.

En la ***situación c,*** el 76% de las aulas eran grandes, 12,7% eran pequeñas y el otro 11,3% eran medianas. En dicha situación se observa que el 35% del total de las aulas se diferencian de la media, este porcentaje es el más alto en las tres situaciones, lo que puede ser debido a que hay un mayor porcentaje de aulas grandes.

Si en la Tabla VI.77 se compara el porcentaje de aulas que se diferencian de la media respecto al porcentaje de aulas que no, se observa que es en las aulas pequeñas donde existe una mayor diferencia entre ambas categorías, sin embargo, en las aulas grandes ocurre lo contrario, ya que los porcentajes son muy similares. Por ejemplo, en las aulas grandes de la *situación a*, hay un 52,1% de las mismas que no se diferencian de la media y un 47,9% que si, mientras que en las aulas pequeñas el 86,3% no se diferencia de la media frente al 13,7% que si. Esta tendencia se mantiene en el resto de situaciones (*b* y *c*). En cuanto a las aulas medianas, aunque no es tan pronunciado como en las aulas pequeñas, también el porcentaje de aulas que no se diferencia de la media dista mucho del porcentaje de aulas que si, así, por ejemplo, en la *situación a*, hay un 65,7% de las aulas medianas que no se diferencian de la media frente al 34,3% que si.

En esta misma tabla también se observa que el porcentaje de aulas de cada tipo no afecta tanto a esta diferencia, ya que, por ejemplo, en la *situación c*, donde entorno al 75% de las aulas son grandes, existe una mayor discrepancia entre el porcentaje de aulas que se diferencia de la media y el que no en cada tipo de aula (pequeña, grande mediana) que en la *situación a,* donde alrededor del 75% de las

aulas eran pequeñas. En esta última situación, por ejemplo, la diferencia de aulas pequeñas que se diferencian de la media respecto a las que no, es de 72,6 puntos porcentuales, en el caso de las aulas medianas de 31,4 y en el caso de aulas grandes de 4,2 puntos, estas diferencias son superiores en la *situación c*, 77,8, 62,5 y 18,2 puntos, respectivamente, pero, sin embargo, es en esta *situación c* donde existe un mayor porcentaje total de aulas que si se diferencian de la media.

Acorde con estos resultados, parece que según se incrementa el tamaño de las aulas, mayor es el porcentaje de las mismas que se diferencian de la media y existe una menor discrepancia entre el porcentaje de aulas que se diferencian de la media respecto al que no, siendo dicha diferencia especialmente pequeña cuando se tiene más de 19 estudiantes por aula. Sin embargo, el porcentaje de aulas de cada tipo no afecta tanto a esta diferencia, pero, si afecta al porcentaje total de aulas que si se diferencian de la media, ya que cuando existe un mayor porcentaje de aulas grandes, existe un mayor porcentaje de aulas totales que se diferencian de la media.

Por lo tanto, en las evaluaciones compuestas por aulas pequeñas y medianas no será posible estimar su crecimiento de manera precisa, ya que un alto porcentaje de las mismas tenderán a no diferenciarse de la media, a no ser que incrementen su tamaño muestral. Así, pues, se recomienda ser cautelosos a la hora de tomar decisiones en este tipo de aulas, invitando a no comparar los resultados en el rendimiento de las mismas, respecto al obtenido en aulas compuestas por un mayor número de estudiantes, ya que, en la mayoría de las aulas pequeñas no será posible estimar su crecimiento real sino se incrementa su tamaño. Además, a la hora de estimar los datos en un modelo multinivel no hay que olvidar el efecto de las estimaciones BLUP, ya que tienden a encoger o suavizar ("*shrunkage*") los resultados estimados de las aulas pequeñas hacia la media global de las escuelas.

Finalmente, se presentan los resultados referidos a las pruebas de Chi cuadrado y de Phi y V de Cramer, que miden la fuerza de asociación entre el tamaño de las aulas y su residuo de crecimiento.

Tabla VI.78. *Pruebas de chi-cuadrado y Phi y V de Cramer*

Aula		*Valor*	*gl*	*Sig.asintótica bilateral*
a	Chi-cuadrado de Pearson	84,160[a]	2	0,000
	Razón de verosimilitudes	75,005	2	0,000
	Asociación lineal por lineal	83,440	1	0,000
	N de casos válidos	882		
b	Chi-cuadrado de Pearson	15,808[b]	2	0,000
	Razón de verosimilitudes	16,318	2	0,000
	Asociación lineal por lineal	15,782	1	0,000
	N de casos válidos	611		
c	Chi-cuadrado de Pearson	24,105[c]	2	0,000
	Razón de verosimilitudes	27,195	2	0,000
	Asociación lineal por lineal	23,168	1	0,000
	N de casos válidos	425		

a. 0 casillas (0,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 22,41.

b. 0 casillas (0,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 22,76.

c. 0 casillas (0,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 16,60.

Aula			**Valor**	**Sig.aproximada**
a	Nominal por nominal	Phi	0,309	0,000
		V de Cramer	0,309	0,000
	Número de casos válidos		882	
b	Nominal por nominal	Phi	0,161	0,000
		V de Cramer	0,161	0,000
	Número de casos válidos		611	
c	Nominal por nominal	Phi	0,238	0,000
		V de Cramer	0,238	0,000
	Número de casos válidos		425	

Tal y como se puede observar en la prueba de chi cuadrado, todos los estadísticos han resultado significativos. En cuanto al valor de χ^2 revela una dependencia (relación) real entre el tamaño de las aulas y su residuo de crecimiento. Para valorar dicha relación se pueden observar los estadísticos Phi y V de Cramer, donde al igual que en la correlación sitúan los valores entre 0 y 1. Estos estadisticos muestran una mayor intensidad o dependencia en la *situación a*, respecto a la *situación b* y *c*, siendo la *situación b*, donde la fuerza de asociación entre el tamaño de las aulas y su residuo de crecimiento es menor (0,16).

CAPÍTULO VII: CONCLUSIONES, DISCUSIÓN, LIMITACIONES Y PROSPECTIVA

En el último capítulo de esta tesis doctoral lo que se pretende es recopilar las conclusiones principales extraídas a partir de los resultados que se han encontrado, y realizar una discusión de las mismas. Esto se hace referido a cada uno de los problemas de investigación planteados y dentro de cada problema se alude a las conclusiones y a la discusión de cada objetivo específico correspondiente. También en este capítulo, se incluyen las limitaciones de la tesis y las posibles líneas de investigación futuras que pueden realizarse como continuación de este trabajo.

VII.1. Problema 1. Estudio de la fiabilidad en las estimaciones de las puntuaciones de los estudiantes y del comportamiento del ERM cuando se modifican elementos esenciales del modelo de crecimiento (varianza y correlación entre el intercepto y la pendiente)

En este primer problema de investigación, donde se pretendía conocer cómo afecta el porcentaje de dispersión fijado en los coeficientes aleatorios de un modelo longitudinal y multinivel, así como la correlación entre los parámetros del intercepto y la pendiente, en la fiabilidad de las estimaciones de las puntuaciones de los estudiantes, y en el comportamiento del ERM, se plantearon dos objetivos específicos. Mediante el primero de ellos se quería garantizar una adecuada recuperación de las características experimentales manipuladas antes de generar los resultados finales, y con el segundo el interés residía en estudiar la fiabilidad de las estimaciones en las puntuaciones de los estudiantes, y el comportamiento del ERM, a través del estudio de los patrones de correlación generados entre los niveles de logro de los estudiantes a lo largo del tiempo.

Las conclusiones a las que se llegaron en cada uno de estos objetivos específicos se presentan a continuación.

VII.1.1. Objetivo específico 1. Garantizar una adecuada recuperación de las condiciones experimentales manipuladas

En cuanto a la recuperación de los parámetros fijados en la generación de la muestra (ver apartado, *VI.1.1.1. Recuperación del tamaño muestral*), en todos los modelos se simula una distribución muestral compuesta por 10.000 estudiantes repartidos en 400 aulas, donde el número mínimo de estudiantes por aula es de 20 y el máximo de 30, siendo las aulas más comunes aquellas que cuentan con 25 estudiantes. Por lo tanto, los parámetros fijados previamente, esto es, el total de estudiantes (10.000), así como, el tamaño mínimo de estudiantes por aula (20 estudiantes), el máximo (30 estudiantes) y la media de estudiantes por aula (25), se han generado correctamente, garantizando una adecuada recuperación de las características muestrales.

Referido al resto de parámetros generados (ver apartado, *VI.1.1.2. Recuperación de los parámetros de varianza y correlación)*, se observa una mayor precisión en la recuperación de los mismos en el nivel de estudiante respecto al de aulas, que puede ser debido a las diferencias existentes en el tamaño muestral de los diferentes niveles, pues mientras que el nivel de aulas está compuesto por 400 unidades, el nivel de estudiante está formado por 10.000 unidades. Asimismo, cuanto mayor es la correlación fijada entre el intercepto y la pendiente, mayor homogeneidad existe entre los valores de una misma variable. Aun así, a pesar de estas pequeñas diferencias, y tal como evidencian todas las pruebas realizadas, en los diferentes modelos generados existe una gran precisión en la recuperación de las diversas condiciones fijadas, lo que garantizó que los resultados finales se generaran con una adecuada validez y fiabilidad.

A la vista de los resultados obtenidos en el apartado VI.1.1.1. y VI.1.1.2, se puede concluir que existe una adecuada recuperación de todas las condiciones experimentales manipuladas en este primer problema de investigación.

VII.1.2. Objetivo específico 2. Estudio de la fiabilidad de las puntuaciones de los estudiantes y del comportamiento del ERM en distintas condiciones experimentales

En el siguiente apartado VII.1.2, se presentan las conclusiones más relevantes de los resultados obtenidos en este segundo objetivo específico. Las conclusiones y la discusión de las mismas, se presentan organizadas haciendo referencia a cada una de las tres situaciones que se han explicado en el apartado *V.3.3.3. (Procedimiento de generación de modelos)*.

VII.1.2.1. Conclusiones y discusión de los modelos de la situación a

Tras los resultados obtenidos en la ***situación a***, en la cual solo existía variabilidad en ciertos coeficientes aleatorios del modelo longitudinal y multinivel, se puede concluir lo siguiente:

A partir de los modelos analizados en esta situación, se puede afirmar que **la precisión con la que se estima las puntuaciones de los estudiantes no se encuentran afectadas por el nivel (nivel de aula o nivel de estudiante) en el que se encuentre varianza**, siempre y cuando la dispersión existente en los diferentes coeficientes aleatorios sea la misma. Es decir, si los datos de dos modelos diferentes cuentan con la misma varianza en el error intrasujeto (por ejemplo, 95% de la varianza total del modelo) pero en uno de ellos solo existe variabilidad a nivel de aula (5%) y en otro en el nivel de estudiantes (5%), siendo los coeficientes en los que hay varianza y el valor de los mismos igual en ambos modelos (modelo 6 y 2, respectivamente), se obtiene una matriz de correlaciones entre las puntuaciones de los estudiantes muy similar, aunque el número de unidades en cada nivel sea diferente.

Sin embargo, aunque dichos modelos proporcionan resultados muy semejantes suponen interpretaciones muy diferentes. En el modelo donde exista varianza en el nivel de estudiante, los valores de correlación entre las puntuaciones se deben sobre todo a las características sistemáticas de los mismos, en el modelo donde hay varianza en el nivel de aula, los resultados se ven afectados especialmente por las características de las aulas, pero no de los estudiantes. **Por lo tanto, resultados semejantes pueden deberse a condiciones muy diferentes (efectos**

de las características del aula o de los estudiantes, o de ambos si existe varianza en los dos niveles), aspecto que hasta el momento nadie había demostrado.

Además, los resultados obtenidos en los modelos simulados en esta *situación a,* evidencian que cuando el rendimiento académico es distinto a lo largo del tiempo (se encuentra variabilidad en la pendiente, independientemente del nivel), aunque el porcentaje de variabilidad sea muy pequeño (por ejemplo, 5%), respecto a la varianza total del modelo (95%), **el patrón de correlación que se observa es un patrón acumulativo, es decir, a medida que se realizan más mediciones se incrementan los valores de correlación, siendo las estimaciones más fiables**, tal y como también demostraron Nesselroade et al. (1980) y, posteriormente Rogosa y Willet (1985).

Por lo comentado anteriormente se deduce que, **en los modelos de crecimiento, no siempre se genera un patrón decreciente (simplex) entre las ocasiones de medida**, tal y como muchos autores (Humphreys, 1960; Joreskog, 1970, Joeskog & Sorborm, 1977 y Wert & Hilton, 1977), consideraban que era lo esperable.

Sólo en algunos modelos de la *situación a*, cuando los niveles de logro iniciales, así como el rendimiento a lo largo del tiempo es diferente entre los estudiantes (existe varianza en el estatus y en la pendiente de un mismo nivel), se observa que, dentro de la primera ocasión de medida, es decir, entre la primera medición y el resto de aplicaciones, el patrón tiende a ser decreciente. Esta tendencia se mantiene dentro de otras mediciones, sobre todo en los modelos de la subsituación 8, donde los estudiantes parten de diferentes niveles de logro promedio y su pendiente de crecimiento también es distinta, siendo la variabilidad encontrada en dichos coeficientes muy superior a la del error intrasujeto. En estos modelos (modelo 8, 8.1 y 8.2), se observó que más allá de la primera ocasión de medida, dentro de una misma medición, el valor de correlación entre dos aplicaciones, es menor cuanta más distancia exista entre las mismas. Así, por ejemplo, la correlación entre la ocasión dos y tres era superior a la correlación entre la ocasión dos y cuatro y, esta última, superior a la correlación entre la ocasión dos y cinco, y así sucesivamente. Por lo tanto, en ciertas ocasiones de medida, de estos modelos se ha encontrado el patrón simplex, al que muchos autores (Balzarini, Macchiavelli & Casanoves, 2005; Garson, 2013; Guttman, 1954; Humphreys, 1960, 1968; Kenny, 1979; Lunneborg & Lunnerborg,

1970; Pardo & Ruiz, 2013; Raudenbush. & Bryk, 1986; 2002; Ronald et al., 2010; Vallejo 1989; Vallejo & Fernández, 1990) aluden cuando se habla del patrón esperable entre las puntuaciones de los sujetos a lo largo del tiempo.

Sin embargo, cuando se encuentra un modelo donde el porcentaje de la varianza del error intrasujeto es muy superior (por ejemplo, 95%), al de la varianza del intercepto (por ejemplo, 2,5%) y de la pendiente (por ejemplo, 2,5%) (modelo 7), dicho patrón decreciente que se había obtenido en los modelos anteriores, dentro de las primeras ocasiones de medida, desaparece y pasa a ser constante e incluso creciente dentro de ciertas mediciones.

En definitiva, el patrón simplex es esperable dentro de algunas ocasiones de medida, cuando la varianza del error intrasujeto es más pequeña (inferior al 35%) que la suma de las varianzas existente en el resto de coeficientes.

Si la varianza del error intrasujeto es muy pequeña (5%), respecto a la varianza de la pendiente (95%) (modelo 3), con pocas ocasiones de medida, como son tres, se obtiene información bastante fiable sobre las puntuaciones de los estudiantes, haciendo que no sea necesario realizar un número elevado de mediciones, ya que, la poca información extra que aportan puede no compensar el coste que supone la realización de las mismas. Por el contrario, a medida que aumenta la variabilidad del error intrasujeto, respecto a la varianza de la pendiente, se va reduciendo el valor de las correlaciones, especialmente, si el número de aplicaciones no es elevado.

Así pues, la fiabilidad de las estimaciones en los modelos en los que existe variabilidad en el error intrasujeto y en la pendiente, depende de la cantidad de dispersión que hay en cada uno de estos coeficientes, a mayor varianza en el error intrasujeto menos fiables son las estimaciones de las puntuaciones de los estudiantes.

En cuanto al efecto de la varianza del intercepto, a la vista de los resultados obtenido se puede concluir que, cuando no existe variabilidad en la pendiente, pero si en el estatus, es decir, cuando los niveles de rendimiento de partida de los estudiantes o de las aulas entre sí, son distintos, pero dicho nivel de rendimiento se mantiene estable a lo largo del tiempo, las correlaciones entre la primera ocasión de medida y el resto de mediciones, son superiores a las que se obtienen cuando los estudiantes o las aulas entre sí, parten del mismo rendimiento inicial pero su nivel de

logro es distinto a lo largo de las diferentes aplicaciones. Sin embargo, en esta última situación donde se encuentra varianza en la pendiente, pero no en la varianza del intercepto, las correlaciones a partir de la primera ocasión de medida son superiores, a las que se obtienen cuando los niveles de logro iniciales de los estudiantes son diferentes pero su ritmo de crecimiento es el mismo. Esto es debido a que **los coeficientes aleatorios del intercepto no son acumulativos, a diferencia de la varianza de la pendiente que va multiplicada por el tiempo, y, por tanto, a medida que se realizan más mediciones se incrementan los valores de correlación entre unas ocasiones y otras.**

En general, la matriz de correlaciones que se genera cuando sólo hay variabilidad en el error intrasujeto y en el intercepto, es una matriz muy homogénea, donde la fiabilidad de las estimaciones a lo largo del tiempo es muy similar, puesto que no hay efecto de la pendiente de crecimiento en las distintas ocasiones de medida, por lo que siempre se estima una puntuación muy parecida. Cuando la dispersión en el estatus es muy elevada (por ejemplo, 95%) en comparación con la varianza del error intrasujeto (por ejemplo, 5%) (modelo 6 y modelo 6.1), las correlaciones entre la primera ocasión de medida y el resto de ocasiones son muy altas, y esos valores se mantienen constantes a lo largo de toda la matriz de correlaciones. Lo contrario ocurre cuando el porcentaje de la varianza del error intrasujeto es muy elevado (por ejemplo, 95%), en comparación con el existente en la varianza del estatus (por ejemplo, 5%) (modelo 5).

En cuanto al ERM, el **patrón constante** entre la primera ocasión de medida y el resto de las aplicaciones, que conlleva a que dicho efecto se mantenga inmutable a lo largo del tiempo, independientemente del número de mediciones que se realicen (Nesselroade et al., 1980), se ha encontrado, sobre todo, en los modelos de esta situación cuando todos los estudiantes y aulas parten del mismo nivel de rendimiento y su crecimiento es constante a lo largo del tiempo, existiendo toda la variabilidad en el error intrasujeto, cuando el rendimiento de los estudiantes o de las aulas es distinto a lo largo de las aplicaciones (existía variabilidad en la pendiente) pero se mantiene constante en la primera ocasión de medida (no hay varianza en el intercepto) o cuando existía variabilidad solo en el estatus y en el error intrasujeto. El **patrón decreciente** entre la primera ocasión de medida y el resto de aplicaciones, que

supone que a medida que se incrementan el número de mediciones el ERM es más marcado (Nesselroade et al., 1980) se encontró cuando había dispersión en el error intrasujeto y en el estatus y en la pendiente de un mismo nivel, sobre todo, si existía mucha diferencia entre la cantidad de varianza del estatus y la de la varianza del error intrasujeto, siendo menor este último coeficiente. Aunque el **patrón creciente,** que indica que a medida que se realizan más mediciones, se incrementa el valor de la correlación entre las puntuaciones de los estudiantes en las distintas ocasiones de medida, pudiendo llegar a desaparecer el ERM si la correlación entre alguna de las mediciones es perfecta (Nesselroade et al., 1980), no se ha observado en ningún modelo de esta situación entre la primera ocasión de medida y el resto de aplicaciones, pero se ha encontrado una correlación perfecta dentro de la primera ocasión de medida cuando no existía variabilidad en el error intrasujeto y toda la variabilidad se encontraba en el estatus (modelo 5.2).

Las **ideas más importantes** a las que se llegan en dicha *situación a,* son las siguientes:

1. El nivel de aula o estudiante, en el que existe variabilidad no afecta a la fiabilidad con la que se estiman las puntuaciones de los estudiantes, siempre que la cantidad de dispersión sea igual en los mismos coeficientes.
2. Se observa un patrón acumulativo entre las puntuaciones de los estudiantes en las diferentes mediciones, cuando su rendimiento o el de las aulas es distinto a lo largo del tiempo.
3. La varianza del estatus a diferencia de la de la pendiente no tiene un efecto acumulativo, por ello, cuando los estudiantes o las aulas parten de distintos niveles de logro, pero el mismos se mantiene constante a lo largo del tiempo se obtiene una matriz de correlaciones muy homogénea
4. No siempre que se trabaje con datos longitudinales es esperable un patrón simplex o decreciente, solo cuando la dispersión del error intrasujeto sea inferior a la del resto de coeficientes.
5. Cuando el porcentaje existente en la varianza de la pendiente es muy elevado en comparación con el del error intrasujeto, se necesitan menos ocasiones de medida para obtener estimaciones fiables del rendimiento de los estudiantes, que en la situación contraria.

6. En cuanto al ERM en los modelos de esta situación, se ha encontrado fundamentalmente un patrón constante y decreciente entre la primera ocasión de medida con el resto de mediciones, por lo tanto, este fenómeno se mantiene constante o se incrementa a lo largo del tiempo. Excepto cuando no hay varianza en el error intrasujeto y toda la variabilidad se debe al intercepto, en este caso el ERM no existe, ya que las correlaciones entre la primera ocasión de medida y el resto de aplicaciones son perfectas.

A la vista de estas conclusiones se realizan las siguientes **sugerencias** de cara a la realización de evaluaciones educativas de calidad.

a) Prácticamente en condiciones irreales (con un residuo muy pequeño, máximo del 5%) se puede considerar que tres medidas son suficientes para realizar un acercamiento al rendimiento real de los sujetos.
b) Generalmente, si los niveles de logro de los estudiantes son diferentes a lo largo del tiempo (existe variabilidad en la pendiente), a medida que se realizan más mediciones, más fiables son las estimaciones de las puntuaciones de los estudiantes.
c) Si el rendimiento de los estudiantes es constante a lo largo del tiempo, la fiabilidad de las estimaciones en las puntuaciones de los estudiantes va a ser siempre muy similar, independientemente del número de pruebas que se realicen, por lo tanto, no habrá apenas diferencia entre realizar tres mediciones o nueve.

VII.1.2.2. Conclusiones y discusión de los modelos de la situación b

A partir de los resultados obtenidos en la ***situación b***, caracterizada porque existía variabilidad en el error intrasujeto, varianza en el intercepto y la pendiente de un mismo nivel y, distintos valores de correlación entre el intercepto y la pendiente de dicho nivel, se llega a las siguientes conclusiones.

Los valores de correlación existentes entre el intercepto y la pendiente de uno de los niveles (aula o estudiantes), afectan a la precisión con la que se estiman las puntuaciones de los estudiantes, aunque no de manera tan marcada como la variabilidad de los coeficientes aleatorios.

Al comparar modelos donde la intensidad de relación entre el intercepto y la pendiente es distinta, pero con una misma direccionalidad, no existen muchas diferencias en los resultados obtenidos. Así, por ejemplo, la matriz de correlaciones generada cuando no existe relación entre el intercepto y la pendiente, y la obtenida cuando dicha relación es baja (0,2) es muy similar. Cuando existe una tendencia más marcada, en la cual, los estudiantes que parten de un nivel de logro alto crecen mucho a lo largo del tiempo y viceversa para los que parten de un nivel de logro bajo (correlación de 0,9), las puntuaciones son más fiables con menos ocasiones de medida que cuando se encuentra un valor de correlación entre el intercepto y la pendiente bajo (0,2), aunque dicha diferencia, en general, no es muy marcada.

Si se comparan dos modelos donde en ambos los estudiantes que parten de un rendimiento bajo incrementan mucho sus niveles de logro a lo largo del tiempo, y, por el contrario, los que parten de un rendimiento alto lo mantienen constante, siendo dicha tendencia en uno de ellos poco pronunciada (se encuentra un valor de correlación entre el intercepto y la pendiente de -0,2) y en el otro más pronunciada (se encuentra un valor de correlación de -0,9), las matrices generadas en estos dos modelos tienden a diferir más entre sí que las obtenidas al comparar modelos donde la relación entre el intercepto y la pendiente es de direccionalidad positiva. Por supuesto, los modelos que más se diferencian entre sí, son aquellos donde existe una relación entre el estatus y la pendiente de intensidad alta, pero en uno de ellos la direccionalidad es positiva y en otro negativa.

Cuando se estudian modelos donde existen diferentes valores de correlación entre el intercepto y la pendiente, las diferencias en la fiabilidad de las estimaciones de las puntuaciones de los estudiantes, entre unos modelos y otros, son más marcadas cuando el número de medidas es pequeño (tres o cuatro), ya que cuando el número es elevado, se sigue observando el efecto acumulativo de la pendiente. Por lo tanto, generalmente, **con muchas ocasiones de medida como son más de cinco, el efecto del valor de correlación encontrado entre el intercepto y la pendiente es poco patente, debido a que la precisión con la que se estima las puntuaciones de los estudiantes es muy similar, tanto con cinco como con nueve mediciones.**

Cuando se encuentran modelos con diferentes valores de correlación entre el intercepto y la pendiente, se ha observado que, se obtienen estimaciones más fiables de las puntuaciones de los estudiantes con un menor número de ocasiones de medida, en aquellos modelos donde el rendimiento del que parten los estudiantes o las aulas, se encuentra muy relacionado con el que alcanzan a lo largo del tiempo. Siendo los estudiantes o las aulas que parten de niveles de logro alto, los que más crecen en las diferentes ocasiones de medida, y, por el contrario, las aulas o los estudiantes que tienen un rendimiento inicial bajo, los que menos incrementan dichos niveles. Es decir, cuando existe una correlación entre el intercepto y la pendiente alta positiva (0,9).

Por el contrario, los modelos que muestran estimaciones del rendimiento menos fiables con un mayor número de mediciones, son los modelos compensatorios, en los cuales los estudiantes o las aulas que parten de niveles de logro alto, su rendimiento se mantiene constante o no crece mucho a lo largo del tiempo**.** Mientras que los estudiantes o las aulas que parten de niveles de logro bajo, su rendimiento crece mucho a lo largo de las diferentes ocasiones de medida, en la mayoría de los casos. Sobre todo, si la fuerza de asociación entre el rendimiento del que parten y el que alcanzan a lo largo del tiempo es alta. Por lo tanto, cuando se encuentra una correlación entre el intercepto y la pendiente baja de direccionalidad negativa (-0,9), las estimaciones son menos fiables, especialmente si el número de mediciones no es elevado.

Tanto el patrón de correlación obtenido entre la primera ocasión de medida con el resto de aplicaciones, como los generados entre las demás mediciones., es el mismo que el obtenido en los modelos donde existían los mismos coeficientes aleatorios, pero no había relación entre el intercepto y la pendiente. Dicho patrón solo se ha visto alterado cuando existe una correlación entre el intercepto y la pendiente alta, de direccionalidad positiva (0,9), pues bajo esta condición la mayoría de los modelos muestran un patrón de correlación entre las diferentes ocasiones de medida que tiende a ser creciente. **En estos modelos donde se obtiene un patrón creciente entre la primera ocasión de medida y el resto de mediciones, dicho patrón creciente se mantiene o se vuelve constante dentro del resto de aplicaciones**. **Sin embargo, cuando existe un valor de correlación entre el intercepto y la**

pendiente bajo positivo (0,2) o bajo y alto negativo (-0,2 y -0,9), si se puede observar un patrón autorregresivo o simplex, entre la primera ocasión de medida y el resto de ocasiones, extensible dentro de otras mediciones más allá de la primera aplicación.

En el único modelo, donde no se ha encontrado un patrón creciente entre la primera ocasión de medida y el resto de ocasiones, así como entre las correlaciones de la ocasión dos y tres con el resto de mediciones, es cuando existe una correlación entre el intercepto y la pendiente de 0,9, y la variabilidad del error intrasujeto es muy pequeña (5%) en comparación con la del intercepto (47,5%) y la pendiente (47,5%) (modelo 12.1). En este caso, el patrón de correlación dentro de las tres primeras ocasiones de medida tiende a ser decreciente.

En cuanto al ERM, el **patrón constante** se ha encontrado en los modelos de esta situación cuando la cantidad de varianza del error intrasujeto era muy elevada (95%), en comparación con la dispersión del intercepto (2,5%) y, además, la correlación entre el intercepto y la pendiente era de 0,2, -0,2 o de -0,9 (modelo 10, modelo 10.2 y modelo 10.3). El **patrón decreciente** se obtuvo cuando existía variabilidad en el error intrasujeto, en la pendiente y en el estatus de un mismo nivel, sin ser la varianza del error intrasujeto muy superior a la varianza del estatus (subsituación 11 y 13) o incluso inferior (subsituación 12 y 14) y, además, la relación entre el intercepto y la pendiente era baja con direccionalidad positiva (0,2), o baja con direccionalidad negativa (-0,2), o alta con direccionalidad negativa (-0,9). Si la correlación era alta, pero con direccionalidad positiva, solo se encontró un patrón decreciente si el porcentaje de la varianza del estatus, era muy elevado en comparación con el porcentaje de la varianza del error intrasujeto. Finalmente, el **patrón creciente** se generó cuando la correlación entre el intercepto y la pendiente era muy alta y de direccionalidad positiva (0,9), a no ser que la varianza del error intrasujeto fuera muy pequeña en comparación con la varianza del estatus.

Las **ideas principales** más importantes que se han obtenido a partir de los modelos generados en esta situación son las siguientes:

1. La relación entre el punto de partida y la pendiente de crecimiento del nivel de estudiante o del nivel de aula, afecta en la fiabilidad con la que se estiman las puntuaciones de los estudiantes, pero no de manera tan marcada como la dispersión existente en los coeficientes aleatorios.
2. Cuando en distintos modelos la fuerza de asociación entre el rendimiento de partida y la pendiente de crecimiento es diferente, pero la direccionalidad de la relación es la misma, la precisión con la que se estiman las puntuaciones de los estudiantes en todos los modelos, es muy parecida.
3. Cuando se comparan modelos donde la direccionalidad entre el punto de partida y la pendiente de crecimiento es distinta, se obtienen estimaciones más precisas con menos ocasiones de medida, en los modelos donde el rendimiento de partida está relacionado con el adquirido a lo largo del tiempo (direccionalidad positiva).
4. El efecto de la relación entre el punto de partida y la pendiente de crecimiento, sobre la precisión con la que se estiman las puntuaciones de los estudiantes a lo largo del tiempo, es más marcado cuando el número de mediciones realizadas es pequeño (tres o cuatro aplicaciones).
5. Los modelos más informativos, con menos ocasiones de medida son aquellos donde la intensidad de relación entre el rendimiento de partida y el crecimiento es alta, de direccionalidad positiva, sin embargo, los menos fiables son aquellos cuya correlación entre el estatus y la pendiente es alta también, pero de direccionalidad negativa (modelo compensatorio).
6. Cuando existe una tendencia muy marcada donde los estudiantes que parten de un nivel de rendimiento alto crecen mucho a lo largo de las aplicaciones, y lo contrario para los que tienen un rendimiento inicial bajo, no se suele observar un patrón simplex, sino una predisposición hacia un patrón creciente dentro de las diferentes mediciones. Por lo que el ERM podría llegar a desaparecer si la correlación entre la primera ocasión de medida y alguna otra fuera perfecta.

Acorde con las conclusiones obtenidas en dicha *situación b*, **las consideraciones** que deberían tener en cuenta los especialistas e interesados en las evaluaciones educativas de calidad son las siguientes:

a) En evaluaciones donde exista una fuerte tendencia en la cual, el rendimiento de partida de los estudiantes o las escuelas no se encuentra muy relacionado con el adquirido a lo largo del tiempo, existiendo una variabilidad equilibrada entre los distintos coeficientes aleatorios (por ejemplo, modelo 13.3), se requieren al menos cinco o seis mediciones para poder obtener estimaciones precisas sobre el rendimiento de los estudiantes (correlaciones superiores a 0,90).Sin embargo, en esta misma situación cuando la direccionalidad entre la relación del estatus y la pendiente es positiva, con tres o cuatro medidas se puede predecir las puntuaciones de los estudiantes en los cursos posteriores, ya que generalmente, se obtienen correlaciones entre las puntuaciones de los estudiantes superiores a 0,90.
b) Por lo tanto, las aulas o escuelas compensatorias que intentan nivelar los niveles de logro de sus estudiantes, es decir, que los que partan de un rendimiento previo bajo crezcan mucho a lo largo del tiempo, y los que partan de un nivel de rendimiento alto se mantengan constantes entre las aplicaciones, requieren de un mayor número de mediciones para estimar con ciertas garantías de fiabilidad el rendimiento de los estudiantes, que aquellas escuelas donde el logro inicial de los estudiantes esta muy relacionado con el adquirido a lo largo del tiempo.

VII.1.2.3. Conclusiones y discusión de los modelos de la situación c

A partir de los resultados generados en los modelos de la ***situación c***, donde existía variabilidad en todos los coeficientes aleatorios, y distintos valores de correlación entre el intercepto y la pendiente del nivel de estudiante y del nivel de aula, se llegan a las siguientes conclusiones.

Acorde con la matriz de correlación obtenida en el modelo 15, que se desarrolló a partir de los valores de varianza encontrados en diferentes investigaciones educativas de carácter multinivel y longitudinal (Castro, et al., 2009;

Navarro; 2013; Raudenbush & Bryk, 2002; Zvoch & Stevens, 2003; Zvoch & Stevens, 2008), se puede concluir **que para poder predecir el rendimiento de los estudiantes con cierta precisión (correlaciones superiores a 0,90), se requieren al menos tres mediciones, que son las ocasiones mínimas para poder desarrollar un modelo de crecimiento. Con cuatro aplicaciones, se obtienen estimaciones muy fiables del rendimiento de los estudiantes, ya que las correlaciones eran en torno a 0,95**.

En el modelo 15, no se ha encontrado mucha diferencia entre la información que proporcionan cinco mediciones (correlación de 0,97) y la obtenida con nueve (correlación de 0,99), por ello en este modelo, al igual que ocurre en otros muchos, **realizar más de cuatro o cinco aplicaciones no es rentable, ya que la información extra que aporta no compensa el coste que conlleva la realización de las mismas.**

En estos modelos de la situación c, caracterizada porque tanto los estudiantes como las aulas entre si parten de diferentes niveles de logro y su rendimiento a lo largo del tiempo es distinto, se observa que si dos modelos de esta situación cuentan con la misma variabilidad en los mismos coeficientes aleatorios, y los mismos valores de correlación entre el intercepto y la pendiente, pero con el orden invertido, por ejemplo, en un modelo la correlación entre el intercepto y la pendiente del nivel de aula es 0,2 y la del nivel de estudiante 0,9, y en otro modelo al revés, es decir, la correlación entre el estatus y la pendiente a nivel de estudiante es de 0,2 y en el nivel de aula de 0,9, **generalmente los valores de correlación entre las puntuaciones de los estudiantes en ambos modelos, son muy similares**. Aunque, en algún caso puntual, se puede encontrar que las correlaciones son ligeramente superiores en el modelo en el cual se encuentra la correlación mayor en el nivel de estudiante, esto puede ser debido a que había más unidades de análisis en este nivel (10.000 estudiantes) que en el nivel tres (400 aulas).

Cuando se compara los resultados obtenidos en un modelo en el que existe solo correlación entre el intercepto y la pendiente en un único nivel, con un modelo en el que además de tener igual valor de correlación en ese mismo nivel (aula o estudiante), existe un valor de correlación entre el intercepto y la pendiente del otro nivel, en general, **la fiabilidad de las estimaciones en ambos modelos es muy similar**. Esto puede ser debido a que, tal y como se ha comentado, el valor de

correlación encontrado entre el intercepto y la pendiente no tiene un efecto muy marcado en la fiabilidad con la que se estiman las puntuaciones de los estudiantes a lo largo del tiempo.

Del mismo modo, cuando se comparan los resultados de modelos en los cuales hay la misma dispersión en los coeficientes aleatorios, pero distintos valores de correlación entre el intercepto y la pendiente del nivel de estudiante y del nivel de aula, si la direccionalidad de los valores de correlación fijados entre el intercepto y la pendiente es la misma en ambos niveles de los distintos modelos, **los resultados de dichos modelos son muy parecidos, aunque la intensidad de los valores de correlación fijados entre el intercepto y la pendiente sean muy distintos.** Por ello, no existe mucha diferencia entre la matriz de correlación obtenida en un modelo donde la correlación entre el intercepto y la pendiente del nivel del aula y del nivel de estudiante es de 0,2 (combinación 1), y la que se obtiene cuando dichos valores de correlación son de 0,9 (combinación 6). Aunque dicha diferencia en los valores de la matriz de correlación es más marcada si estos valores tienen una direccionalidad negativa (combinación 11 y combinación 16, respectivamente).

De todas las combinaciones de la *situación c*, es **la combinación 6, la que muestra estimaciones más fiables con un menor número de ocasiones de medida**, ya que el rendimiento de partida de las aulas y de los estudiantes estaba muy relacionado con los niveles de logro adquiridos a lo largo del tiempo, (la correlación entre el intercepto y la pendiente del nivel de estudiante y del nivel de aula era de 0,9).

Cuando en un modelo se encuentran estos valores de correlación entre el estatus y la pendiente en el nivel de aula y de estudiante (0,9), existiendo una variabilidad equilibrada entre los diferentes coeficientes aleatorios (subsituación 17), con tres o cuatro mediciones se obtienen correlaciones superiores a 0,90 entre las puntuaciones de los estudiantes, en las diferentes ocasiones de medida. Por el contrario, **la combinación que muestra menos precisión en las estimaciones de los niveles de logro de los estudiantes es la 16, ya que, la correlación entre el intercepto y la pendiente del nivel de aula y del nivel de estudiante era de -0,9**. En el modelo donde la variabilidad de los coeficientes aleatorios estan basados en la realidad educativa, y cuentan con una correlación entre el intercepto y la pendiente

de -0,9 en el nivel de aula y de estudiantes (combinación 16, de la subsituación 17), se requieren entre 6 y 7 mediciones para obtener correlaciones fiables, superiores a 0,90.

En cuanto al ERM, en esta situación se encontró un **patrón constante** cuando el porcentaje de la varianza del error intrasujeto, suponía un 95% de la varianza total del modelo, y el 5% restante, se encontraba repartido por igual en el resto de coeficientes aleatorios (subsituación 18). Con estos porcentajes de varianza, el ERM se mantiene constante en todas las combinaciones, a excepción de la combinación 6, en la cual, la correlación entre el intercepto y la pendiente del nivel de estudiante y del nivel de aula era de intensidad alta y direccionalidad positiva (0,9), en este caso, se observó una cierta tendencia hacia un patrón creciente. Respecto al **patrón decreciente**, se encontró en el modelo 15, que se desarrolló a partir de los resultados obtenidos en distintas investigaciones educativas (Castro et al., 2009; Navarro; 2013; Raudenbush & Bryk, 2002; Zvoch & Stevens, 2003; Zvoch & Stevens, 2008), en el modelo generado a partir del modelo original de Raudenbush y Bryk (2002) (modelo 16), en todas las combinaciones del modelo 17 y en el modelo 19. En cuanto al **patrón creciente,** dentro de esta situación apareció en la combinación 6 de la subsituación 18, donde la correlación entre el intercepto y la pendiente del nivel de aula y del nivel de estudiante era de 0,9.

Las **ideas claves** de esta situación se resumen a continuación:

1. El nivel (estudiante o aula) donde se encuentra un mayor valor de correlación entre el intercepto y la pendiente, no afecta de manera marcada a la fiabilidad con la que se estiman las puntuaciones de los estudiantes.
2. No existe apenas diferencia en la precisión con la que se estiman las puntuaciones de los estudiantes, cuando existe solo relación entre el rendimiento de partida y el obtenido a lo largo del tiempo en uno de los niveles (aula o estudiante), que cuando existe en ambos (aula y estudiante).
3. Si en dos modelos existen distintos valores de correlación entre el intercepto y la pendiente del nivel de aula y del nivel de estudiante, pero la direccionalidad es la misma, la fiabilidad con la que se estiman las puntuaciones de los estudiantes es muy similar en dichos modelos. Por lo

tanto, parece que la intensidad de relación entre el intercepto y la pendiente no afecta mucho a la precisión con la que se estiman las puntuaciones de los estudiantes a lo largo del tiempo.

4. Cuando existe relación entre el intercepto y la pendiente en el nivel de aula y en el nivel de estudiante, los modelos que aportan mayor información con un menor número de ocasiones de medida, son aquellos donde tanto el rendimiento de partida de los estudiantes como el de las aulas, se encuentra muy relacionado con su pendiente de crecimiento, en estos modelos con tres o cuatro aplicaciones, la precisión con la que se estiman las puntuaciones de los estudiantes es alta. Por el contrario, cuando existe una correlación entre el intercepto y la pendiente a nivel de aula y a nivel de estudiantes de intensidad alta y direccionalidad negativa (-0,9), se requieren al menos entre seis o siete mediciones, para obtener estimaciones precisas del rendimiento de los estudiantes.
5. Generalmente en las distintas combinaciones de los modelos de la *situación c*, las correlaciones entre la primera ocasión de medida y el resto son decrecientes, por lo tanto, el ERM se incrementa a medida que se realizan más aplicaciones.

A partir de los resultados obtenidos en dicha *situación c*, pero, sobre todo, de los resultados del modelo 15, desarrollado a partir de otras investigaciones, se realizan las siguientes **recomendaciones**:

a) Acorde con los resultados del modelo 15, con tres aplicaciones se puede predecir con precisión el rendimiento de los estudiantes, ya que las correlaciones son superiores a 0,90, pero con cuatro ocasiones de medida las correlaciones son superiores a 0,95.

b) A veces realizar un número muy elevado de mediciones, como es más de cuatro, no es rentable, ya que no aporta mucha información y supone mucho coste. La realización de más de cuatro aplicaciones es sobre todo necesaria cuando se trabajan con datos donde el rendimiento de partida no esta muy relacionado con el crecimiento. En estos casos, especialmente cuando existe una correlación entre el intercepto y la

pendiente del nivel de aula y del nivel de estudiantes de intensidad alta, se requieren al menos seis o siete mediciones para poder estimar el rendimiento de los estudiantes de manera precisa (correlaciones superiores a 0,90).

VII.1.2.4. Conclusiones generales de los modelos de la situación a, b y c

Como **conclusiones generales,** obtenidas a partir de todos los modelos simulados en la *situación a, b* y *c,* se pueden considerar las siguientes:

1. **El factor que más influye en la fiabilidad de las estimaciones de las puntuaciones de los estudiantes es la diferencia de dispersión existente entre el error intrasujeto y la pendiente**. Obteniéndose mayor información sobre las puntuaciones de los estudiantes con menos ocasiones de medida, cuanto mayor sea la varianza de la pendiente, respecto a la varianza del error intrasujeto. Independientemente de si la dispersión de la pendiente se encuentra en el nivel de estudiante o en el nivel de aula.
2. **La varianza del error intrasujeto, respecto a la dispersión del resto de coeficientes aleatorios juega un papel muy importante en la precisión con la que se estiman los niveles de logro de los estudiantes a lo largo del tiempo**. Cuando toda la varianza se acumula en el error intrasujeto, considerando que tanto los estudiantes que van a la misma aula, como las aulas entre si, parten del mismo rendimiento promedio y su crecimiento es constante a lo largo del tiempo, las correlaciones entre las puntuaciones de los estudiantes en las distintas ocasiones de medida son nulas, ya que, no existe variabilidad ni en el estatus ni en el intercepto.
3. **Cuando no existe dispersión en el error intrasujeto se obtiene un patrón de correlación perfecto entre las puntuaciones de los estudiantes**, tanto si toda la variabilidad se encuentra en el intercepto, es decir las aulas o los estudiantes parten de distintos niveles de rendimiento previo, como si los estudiantes o las aulas crecen a distinto ritmo (existe varianza en la pendiente). En esta última situación, se genera un patrón de correlación perfecto, excepto entre las correlaciones de la primera ocasión de medida y el resto de mediciones, ya que, el nivel de rendimiento de partida es el mismo en todos los estudiantes y en todas las aulas.

4. Si la varianza del error intrasujeto es inferior a la del resto de coeficientes aleatorios, se obtiene mayor información sobre el rendimiento de los estudiantes que en la situación contraria, especialmente cuando el número de mediciones no es elevado. De esta manera, **si la varianza del error intrasujeto es elevada (95%) respecto al valor del resto de coeficientes aleatorios, ni con 9 mediciones se consigue estimar con precisión las puntuaciones de los estudiantes (correlaciones inferiores a 0,90). Sin embargo, si la varianza del error es pequeña (5%) con tres medidas ya se obtienen correlaciones entre las puntuaciones de los estudiantes superiores a 0,90.**
5. **Los valores de correlación existentes entre el intercepto y la pendiente, no tienen un efecto muy marcado en la fiabilidad con la que se estiman las puntuaciones de los estudiantes a lo largo del tiempo**, pero la direccionalidad, y la intensidad de relación entre el estatus y la pendiente supone interpretaciones muy diferentes entre la relación del rendimiento de partida y el que se alcanza a lo largo del tiempo.
6. Otra de las aportaciones más importante de esta tesis, obtenida a partir de los resultados generados en las distintas simulaciones, es que, acorde a los patrones de correlación encontrados, **es un error considerar que cuando se trabaja con medidas repetidas cabe esperar únicamente un patrón simplex o autorregresivo,** tal y como, apuntan muchos autores (Garson, 2013; Pardo & Ruiz, 2013; Raudenbush & Bryk, 2002; Ronald et al., 2010, etc.). Este patrón aparece de manera marcada, dentro de diversas ocasiones de medida, cuando tanto el rendimiento de partida, como el adquirido a lo largo del tiempo es distinto (existe dispersión en el intercepto y la pendiente del nivel de aula, o del nivel de estudiante o de ambos) y, además, existe mucha más variabilidad en dichos coeficientes respecto a la varianza del error intrasujeto. A medida que se igualan los porcentajes de varianza del estatus y la pendiente, con el de la varianza del error intrasujeto, dicho patrón se atenúa, llegando a desaparecer cuando la varianza del error intrasujeto es muy superior a la dispersión total de la suma de la varianza del intercepto y la pendiente, a no ser que exista un valor de correlación entre el intercepto y la pendiente que sea negativo.

Cuando los estudiantes y/o aulas, que parten de un nivel de logro alto se mantienen constantes a lo largo del tiempo o no crecen mucho, pero sí se incrementa mucho el rendimiento de los estudiantes y/o las aulas que parten de bajos niveles de logro, es decir, existe una relación compensatoria entre el intercepto y la pendiente, también se puede encontrar un patrón simplex entre ciertas ocasiones de medida.

7. Aunque en determinados modelos, **dentro de algunas ocasiones de medida (por ejemplo, las correlaciones entre la primera ocasión de medida con el resto de aplicaciones) es posible encontrar un patrón decreciente, en casi todas las situaciones simuladas (incluido en los modelos que presentan un patrón autorregresivo dentro de alguna ocasión de medida)**, **se observa que según se añaden más mediciones, entre ellas se produce un incremento de los valores de correlación (patrón creciente).** Así, por ejemplo, aunque puede ocurrir que en algunos modelos la correlación entre la primera y segunda medición sea superior a la correlación entre la primera ocasión con la tercera, y a su vez esta correlación superior a la de la primera ocasión con la cuarta, en todos los modelos donde los estudiantes o las aulas tienen distinto rendimiento a lo largo del tiempo, la correlación entre la ocasión una y dos es inferior a la correlación entre la ocasión dos y tres, y esta inferior a la correlación entre la ocasión tres y cuatro y así sucesivamente, a pesar de existir la misma distancia temporal entre todas ellas, pues existe un efecto acumulativo de la pendiente de crecimiento.
8. **Que a medida que se realicen más aplicaciones, el patrón entre las puntuaciones de los estudiantes a lo largo del tiempo sea creciente, indica que hay efecto de las características de las aulas y/o de las características de los estudiantes sobre el rendimiento**. Si el patrón fuera decreciente supondría, que no existe efecto alguno de las características de los estudiantes y/o de las aulas, sobre los niveles de logro. Representaría una situación semejante a la obtenida cuando los estudiantes cambian continuamente de aula o de escuela en cada medición y, los efectos de las mismas no estuvieran relacionados entre sí. De esta manera, la matriz autorregresiva o simplex es esperable a medida que se añaden aplicaciones, en el caso de que no exista intervención de los centros educativos que influyan

en los procesos de aprendizaje de los estudiantes, y esa ausencia de intervención no se vea suplida por el propio progreso académico del estudiante.

9. **El modelo más fiable con un menor número de ocasiones de medida (3 aplicaciones) es aquel donde hay poca variabilidad en el error intrasujeto, en comparación con la de la pendiente de crecimiento, y existe una relación entre el intercepto y la pendiente de ambos niveles, alta de direccionalidad positiva.**
10. Aunque Raudenbush y Bryk (2002), consideran que ya con dos puntos temporales se pueden estudiar ganancias entre una aplicación y otra, y con tres, es posible la estimación de una curva de crecimiento, cuando se pretende garantizar la fiabilidad de las estimaciones en las puntuaciones de los estudiantes, el número de ocasiones de medidas necesarias, dependen, sobre todo, de la diferencia entre la variabilidad del error intrasujeto y la pendiente de crecimiento, ya que, cuando el porcentaje de varianza del error intrasujeto es muy elevado, el efecto de la varianza de la pendiente se observa a largo plazo. Sin embargo, **cuando se trabaja con valores propios de la realidad educativa (modelo 15) donde no existe una diferencia muy marcada entre ambos porcentajes de varianza, con cuatro mediciones, se suelen obtener estimaciones bastante fiables**. Pues tal y como se demostró en el modelo 15 las correlaciones entre las puntuaciones de los estudiantes con cuatro ocasiones de medida son de 0,95. Por lo tanto, los resultados encontrados se encuentran en sintonía con las recomendaciones realizadas por Gray et al., (1995) sobre la realización de entre tres y cinco mediciones en los diseños longitudinales para poder conocer el rendimiento de los estudiantes. Asimismo,. las evaluaciones realizadas por la NWEA, que se basan en cinco ocasiones de medida, teniendo en cuenta lo estudiado anteriormente, parece que están realizando evaluaciones muy fiables, mediante las cuales es posible estimar el rendimiento de los estudiantes con precisión.
11. En cuanto el ERM entre la primera ocasión de medida y el resto de aplicaciones, en **la mayoría de los modelos simulados este fenómeno se mantiene constante o se incrementa a lo largo de las diferentes ocasiones de medida**, tan sólo tiende a desaparecer (patrón creciente) cuando la

relación entre el intercepto y la pendiente tiene una fuerza de asociación alta, de direccionalidad positiva (0,9).

Como posibles **recomendaciones generales** que se pueden plantear a partir de las conclusiones obtenidas en este primer problema de investigación, con el fin de mejorar la calidad de las evaluaciones educativas se sugieren las siguientes:

a) Cuando en una evaluación educativa se observe que existe poca variabilidad en el error intrasujeto, en comparación con la varianza de la pendiente de crecimiento, existiendo, además, una relación positiva entre el rendimiento de partida y el alcanzado a lo largo del tiempo en ambos niveles, se encuentra usted ante un modelo donde con tres ocasiones de medida ya puede predecir el rendimiento de los estudiantes de manera muy fiable.
b) Se recomienda, en general, la realización de al menos cuatro aplicaciones en las evaluaciones educativas, en las cuales se sigan a los mismos estudiantes, con la finalidad de poder predecir su rendimiento de la manera más precisa posible. Se es consciente de que no siempre es posible la realización de las mismas, entre otros motivos, por el coste que suponen. En tal caso, se aconseja tener en cuenta las limitaciones que esto puede suponer, sobre todo, si se engloban dentro de las evaluaciones de alto impacto, en las cuales se penaliza a las escuelas que no llegan a unos niveles de rendimiento mínimo establecidos. Por este motivo, y por otros muchos, relacionados con la complejidad que supone la medida de los niveles de logro, se invita a la realización de evaluaciones con carácter diagnóstico (evaluaciones de bajo impacto), con el fin de introducir mejoras en los sistemas educativos, acorde a sus limitaciones y necesidades.
c) La evaluación educativa, debe ser entendida como una herramienta que permita potenciar las fortalezas de los sistemas y centros educativos, así, como una ayuda para superar las dificultades, siempre con fines de mejora. Nunca debe ser contemplada como un medio de castigo donde se sancione a las escuelas por sus limitaciones. El centro educativo no es una empresa donde solo importa el rendimiento final, en la escuela lo importante es de donde se parte y hasta donde se llega, dentro de un contexto específico. Lo

fundamental es el cambio que realizan los estudiantes y los colegios a lo largo del tiempo, aunque no todos los estudiantes lleguen a un mismo nivel, pues tampoco todos parten del mismo punto. En este sentido, son las evaluaciones longitudinales que siguen a los mismos estudiantes a lo largo del tiempo, las que ofrecen mayor información sobre el progreso, tanto de las escuelas, como de los estudiantes, y, por tanto, las que van a permitir la realización de programas que mejoren la calidad de la educación de forma rigurosa.

d) Las evaluaciones educativas en las cuales se siguen a los mismos sujetos a lo largo de más de dos ocasiones de medida, no deben preocuparse en sus resultados finales por el ERM, ya que este fenómeno solo se produce entre la primera ocasión de medida con el resto de mediciones. Sin embargo, las evaluaciones basadas en los modelos de ganancia (dos ocasiones) o bien en modelos donde se seleccionan a nuevos sujetos a lo largo del tiempo, el comportamiento del ERM depende del patrón de correlación encontrado entre las distintas aplicaciones. Si el patrón es constante el ERM se mantendrá inmutable en las diferentes ocasiones de medida, si es decreciente, el ERM se incrementará a lo largo del tiempo, y si es creciente puede llegar a desaparecer si se obtiene una correlación perfecta entre la primera medición y alguna otra ocasión de medida.

VII.1.2.4. Conclusiones del estudio del rendimiento de los estudiantes agrupados en los distintos quintiles

Antes de terminar el apartado de conclusión y discusión, referido al primer problema de investigación, se cree conveniente señalar las aportaciones más importantes obtenidas al intentar dar respuesta a la pregunta final, planteada como colofón a dicho problema de investigación: **¿qué aulas potencian más el rendimiento de los estudiantes agrupados en los distintos quintiles?**

Para dar respuesta a esta pregunta se utilizaron los cuatro modelos de la subsituación 14, donde el error intrasujeto se fijó a un 5%, no existía variabilidad en la pendiente de crecimiento de los estudiantes y la varianza del intercepto de ese mismo nivel se fijó a un 15%, la del nivel de aula a un 55% y su pendiente de crecimiento a un 25%. En cuanto a la correlación entre el intercepto y la pendiente

del nivel de aula, en el modelo 14 se fijó a 0,2, en el modelo 14.1, a 0,9, en el modelo 14.2, a -0,2 y, en el modelo 14.3, a -0,9.

De acuerdo con los resultados obtenidos en las aulas de estos modelos agrupados en los cinco quintiles, si se pudiera y/o se tuviera que seleccionar a un aula situada dentro del **primer quintil**, lo ideal sería seleccionar aulas en las que el rendimiento de partida no estuviera muy relacionado con el adquirido a lo largo del tiempo, es decir, aulas compensatorias, donde las que parten de un rendimiento bajo, incrementan mucho el mismo a lo largo del tiempo y, lo contrario, para las aulas con un nivel de logro inicial alto **(aulas del modelo 14.3)**. La elección de dichas aulas en el quintil uno, se debe a que los resultados muestran que en el modelo 14.3, independientemente del tipo de aula (pendiente baja, media o alta) y del quintil en el que se sitúen los estudiantes, todos ellos obtienen una tasa de crecimiento superior a la obtenida en el resto de modelos. Aunque, es en este modelo 14.3, de los modelos de la subsituación 14, el que más mediciones requería para obtener estimaciones fiables del logro de los estudiantes, concretamente con cuatro- cinco se obtenía un valor de correlación de 0,94, y cuando se realizaban cinco o seis medidas, su valor se incrementaba a 0,97.

Entre las aulas del **segundo quintil**, cuando el residuo de la pendiente (valor añadido) es bajo, se obtienen mejores resultados también en las aulas donde el rendimiento de partida no esta muy relacionado con el adquirido a lo largo del tiempo **(modelo 14.3)**, pero cuando el residuo de la pendiente es medio o alto, los resultados son mejores en aquellas aulas donde los niveles de logro obtenidos en la primera ocasión de medida no se relacionan con los obtenidos en el resto de mediciones, sin que dicha tendencia sea muy marcada **(modelo 14).** En este modelo con tres- cuatro mediciones se obtenía un valor de correlación superior a 0,97.

En cuanto a las aulas del **tercer quintil**, si las mismas cuentan con un residuo en la pendiente bajo, los estudiantes muestran una mayor tasa de crecimiento si asisten a aulas donde existe una ligera tendencia en la que los niveles de logro de partida no guardan mucha relación con el adquirido a lo largo del tiempo **(modelo 14.2)**. En este modelo, con tres, cuatro medidas, se obtuvo un valor de correlación de 0,95. Pero cuando el residuo de la pendiente es medio o alto, la ganancia es superior si los estudiantes asisten a aulas donde existe una ligera tendencia en la cual el

rendimiento de partida esta relacionado con el logro adquirido a lo largo de las ocasiones de medida **(modelo 14)**.

Dentro de las aulas situadas en el **cuarto quintil**, en las aulas con un residuo en la pendiente bajo, los estudiantes incrementan más su rendimiento cuando acuden a aulas donde existe una ligera tendencia en la que los niveles de logro de partida no se relacionan mucho con su pendiente de crecimiento **(modelo 14.2)**, pero cuando el residuo de la pendiente de crecimiento es de intensidad media, es mejor que asistan a aulas donde el rendimiento de partida está relacionado con el rendimiento adquirido a lo largo del tiempo, siendo dicha tendencia muy marcada **(modelo 14.1)** o aulas donde existe una ligera tendencia en la cual los niveles de logro iniciales no se encuentran muy relacionados con los obtenidos a lo largo del resto de aplicaciones **(modelo 14.2)**. Si el residuo de la pendiente es alto, los estudiantes obtienen una ganancia superior cuando acuden a aulas que cuentan con una relación entre el intercepto y la pendiente positiva de intensidad baja **(modelo 14).**

Finalmente, en cuanto a las aulas del **quinto quintil**, los estudiantes rinden más, cuando el residuo de la pendiente es bajo o medio en aulas donde existe una ligera tendencia en la que los niveles de logro de partida no guardan mucha relación con el alcanzado a lo largo del tiempo **(modelo 14.2)**, pero cuando el residuo de la pendiente es alto, no existe mucha diferencia entre que asistan a aulas donde el rendimiento de partida está relacionado con el rendimiento adquirido a lo largo del tiempo, sin que dicha tendencia sea muy marcada **(modelo 14)** y asistir a aulas con esa misma tendencia, pero de manera más marcada **(modelo 14.1)**. Sin embargo, ambos modelos difieren en el número de medidas necesarias para obtener estimaciones precisas sobre el rendimiento de los estudiantes, ya que, recuérdese que el modelo 14.1, requiere de menos medidas que el modelo 14. En este último modelo, con tres-cuatro medidas se obtenían correlaciones de 0,97. Este mismo valor se obtuvo en el modelo 14.1 con dos-tres ocasiones de medida.

A la vista de estas conclusiones, las principales **ideas claves** son las siguientes:

1. Aunque, en general, el **modelo 14**, donde la correlación entre el intercepto y la pendiente del nivel de aula es baja positiva (0,2), es el modelo que menos perjudica al rendimiento de los estudiantes, ya que siempre dentro de alguna de las aulas situadas en los distintos quintiles hay estudiantes que obtienen una ganancia positiva entre la primera ocasión de medida y la última, son las aulas del quintil cinco del **modelo 14.1** y del **modelo 14.2,** junto con las aulas del quintil uno del **modelo 14.3**, las que más potencian el crecimiento de los estudiantes de todos los quintiles.
2. Por el contrario, es en el modelo donde el rendimiento de partida está relacionado con el rendimiento adquirido a lo largo del tiempo, sin que dicha tendencia sea muy marcada **(modelo 14.1),** donde en general las aulas de los diferentes quintiles muestran un peor rendimiento. Siendo los estudiantes que asisten a aulas del **primer quintil** de este modelo, los que en conjunto peor tasa de rendimiento muestran entre la primera ocasión de medida y la última, ya que todos los estudiantes situados en los distintos quintiles de estas aulas, sufren una pérdida en sus niveles de logro entre la primera aplicación y la última, pero también en las aulas situadas en el primer quintil donde la correlación entre el intercepto y la pendiente es baja de direccionalidad negativa **(modelo 14.2)** y en las aulas del quintil cinco del modelo donde la correlación entre el intercepto y la pendiente del nivel de estudiante es también negativa, pero de intensidad alta **(modelo 14.3)** se ha observado que los estudiantes de todos los quintiles experimentan una pérdida importante en sus niveles de logro a lo largo del tiempo.
3. Aunque dentro de una misma aula, no existe mucha diferencia entre la tasa de crecimiento de los estudiantes organizados en los distintos quintiles, sí existe una gran diferencia en la tasa de crecimiento de los estudiantes que asisten a aulas donde la relación entre el intercepto y la pendiente difiere en direccionalidad e intensidad. Por lo tanto, **dicho valor de correlación tiene un impacto muy importante sobre la tasa de crecimiento de los estudiantes entre la primera ocasión de medida y el resto de ocasiones.** Sin embargo, no era un factor que afectara mucho en el número de medidas necesarias para realizar estimaciones fiables del rendimiento.

A partir de todas estas ideas se realizan las siguientes **recomendaciones**:

a) Debido a la importancia que tiene la intensidad de relación y la direccionalidad existente entre el rendimiento de partida de las aulas y su pendiente de crecimiento, en la tasa de cambio de los estudiantes, si fuera posible, sería adecuado estudiar la misma en las diferentes aulas de los centros educativos, con el fin de que los estudiantes puedan asistir a aquellas aulas que más potencien su rendimiento académico.
b) Se recomienda potenciar en los sistemas educativos las aulas compensatorias, ya que, además de nivelar el rendimiento de los estudiantes, haciendo que los estudiantes cuyo rendimiento inicial es bajo crezca mucho a lo largo del tiempo, y los que parten de niveles de logro altos se mantengan constante, son las aulas donde los estudiantes más incrementan su tasa de crecimiento y menos pérdida experimentan entre la primera ocasión de medida y la última.
c) Se recomienda encarecidamente estudiar y trabajar con aquellas aulas situadas en los diferentes quintiles, donde la tasa de crecimiento de los estudiantes pertenecientes a los distintos quintiles, decrece a medida que se realizan más aplicaciones.

VII.2. Problema 2. Influencia del tamaño muestral de las aulas en la estimación de su pendiente de crecimiento

En este segundo problema de investigación, donde se pretendía conocer cómo afecta el tamaño de las aulas en la estimación de su pendiente de crecimiento, se plantearon dos objetivos específicos. Mediante el primero de ellos se quería garantizar una adecuada recuperación de las características experimentales manipuladas, antes de generar los resultados finales, y en el segundo el interés residía en estudiar el efecto del tamaño de las aulas sobre la precisión con la que se estima la pendiente de crecimiento de las mismas, en tres circunstancias distintas, donde se fijan diferentes tamaños de aulas pequeñas, medianas y grandes.

Las conclusiones a las que se llegaron en cada uno de estos objetivos específicos se presentan a continuación:

VII.2.1. Objetivo específico 1. Garantizar una adecuada recuperación de las condiciones experimentales manipuladas

A la vista de los resultados obtenidos en el *apartado VI.2.1.* referido al primer objetivo específico de este segundo problema de investigación, se puede concluir que las condiciones experimentales fijadas, dentro de cada una de las situaciones (*situación a, b y c*) o modelos planteados (modelo 15.a, modelo 15.b y modelo 15.c), se generan correctamente, ya que en todas las situaciones se han simulado un porcentaje de aulas próximo al definido previamente, con el número mínimo y máximo especificado en cada tipo de aula (pequeña, mediana y grande). Tan sólo en las aulas grandes, donde se pretendía generar un máximo de 30 estudiantes por aula, se generó un máximo de 31 en la *situación a* y *c,* y en la *situación b,* un máximo de 29 estudiantes, lo que puede ser debido a que las distribuciones muestrales se generan al azar.

Esto no supone un problema para poder dar respuesta al segundo objetivo específico de investigación, ya que, en las aulas grandes lo importante no es el tamaño máximo generado, sino que el mínimo sea de más de 19 estudiantes, pues muchos autores (Bryk & Raudenbush, 2002; Goldstein, 1997; Raudenbush & Bryk, 1986), consideran que 20, son las unidades mínimas para poder llevar a cabo un modelo multinivel. Además, en cuanto al tamaño máximo, acorde a la legislación educativa (LOMCE, 2016) el número máximo de estudiantes en Educación Secundaria Obligatoria es de 30 por aula, pudiéndose contemplar una subida de hasta un 10% de la ratio máxima establecida, para hacer frente a las necesidades inmediatas de escolarización.

Por todo lo comentado anteriormente en este apartado, se puede concluir que existe una adecuada recuperación de todas las condiciones muestrales manipuladas para dar respuesta a este segundo problema de investigación

VII.2.2. Objetivo específico 2. Estudio del efecto del tamaño muestral de las aulas en su pendiente decrecimiento

A partir de los resultados obtenidos (*apartado VI.3.1.*) en cada una de las situaciones generadas, se pueden realizar las siguientes conclusiones:

En la ***situación a,*** donde el 74,5 % de las aulas eran pequeñas (menos de 11 estudiantes por aula), el 12,2% medianas (menos de 20 estudiantes por aula) y el 13,3% grandes (más de 19 estudiantes), **se observó que en las aulas pequeñas existía una gran diferencia entre el porcentaje de aulas que no se diferenciaban de la media (86,3%) y el que si (13,7%), esta diferencia entre las aulas que no se diferencian de la media y las que si, no era tan marcada en las aulas medianas (31,4 puntos porcentuales de diferencia), y aún menos marcada lo era en las aulas grandes (4,2 puntos porcentuales de diferencia).**

En la ***situación b*****,** donde se reduce a un 12,8% las aulas pequeñas, y se incrementa a un 74,8% las aulas medianas, se observó nuevamente **la tendencia anterior, donde a medida que se incrementa el número de estudiantes por aula, menos diferencias se observan entre el porcentaje de aulas que se diferencian de la media y el que no.**

En la ***situación c*****,** donde el 11,3% de aulas eran medianas, el 12,7% pequeñas y el 76% de las aulas eran grandes se destaca que, **aunque dentro de cada tipo de aula (pequeña, mediana y grande) la diferencia entre las aulas que no se diferencian de la media y las que si, no es menor que en el resto de situaciones (*a* y *b*), si se observó que es en esta situación donde existe un mayor porcentaje de aulas totales que se diferencian de la media, concretamente un 35%, mientras que en la *situación a* este porcentaje de aulas era del 20,7% y del 30% en la *situación c.***

Por lo tanto, las **ideas principales** que se extraen de este segundo objetivo específico son:

1. Cuanto mayor sea el tamaño de las aulas, mayor es el porcentaje de las mismas que se diferencian de la media.
2. Según se incrementa el tamaño de las aulas, existe una menor discrepancia entre el porcentaje de aulas que se diferencian de la media respecto al que no, siendo dicha diferencia especialmente pequeña cuando se tiene más de 19 estudiantes por aula, por lo tanto, la recomendación de Bryk y Raudenbush (2002), Goldstein (1997), Raudenbush y Bryk (1986) y Snijders y Bosker (2012) de que almenos se trabaje con 20 estudiantes por aula, para que no exista sesgo en las estimaciones de la pendiente de crecimiento de las mismas, es muy adecuada.
3. El porcentaje de aulas de cada tipo (pequeña, mediana o grande) no afecta a que exista una menor discrepancia entre el porcentaje de aulas que se diferencian de la media.
4. Sin embargo, se produce un salto sustancial entre el número total de aulas que se diferenciaban de la media cuando alrededor del 75% de las aulas eran pequeñas (*situación a*) que cuando ese porcentaje eran aulas grandes (*situación c*), en esta última situación se diferenciaban de la media alrededor de un 15% más de aulas. Esta diferencia no es tan marcada cuando se compara la *situación b* y *c,* siendo un 5% superior las aulas que se diferenciaban de la media en la *situación c*.

Por todo lo comentado en este segundo objetivo específico se realizan las siguientes r**ecomendaciones** respecto a la importancia del tamaño muestral en la realización de evaluaciones educativas:

a) Se recomienda realizar evaluaciones compuestas por más de 19 estudiantes por aula, ya que es cuando existe un mayor porcentaje de aulas que se diferencian de la media, y es cuando menos discrepancia existe entre el porcentaje de aulas que si se diferencia de la media y el que no. Si no es posible, se aconseja tener en cuenta las repercusiones que puede tener trabajar con este tipo de aulas a la hora de estimar su pendiente de crecimiento.

b) No se recomienda tomar decisiones, especialmente de alto impacto (*hight-stakes*), basadas en el crecimiento de aulas compuestas por un bajo número de estudiantes (menos de 20 estudiantes), pues, en muchas de ellas no será posible conocer su rendimiento real, ya que tenderán a no diferenciarse de la media debido a su error de estimación.

c) Se recomienda ser muy cautelosos a la hora de realizar rankings de posicionamiento de los centros educativos, basados en las estimaciones de la pendiente de crecimiento de este tipo de aulas pequeñas y medianas, ya que dichas pendientes de crecimiento se estiman con tanto error que hace que la mayoría no se diferencien de la media.

VII.3. Limitaciones y prospectiva de futuro

En primer lugar, hay que decir que todas las investigaciones, incluso las más cuidadosas tienen limitaciones.

Concretamente, en esta tesis, la realización de una simulación de esta envergadura, con tres niveles de anidamiento, 1.000 réplicas y una gran variedad de combinaciones, supone mucho tiempo de computación para poder desarrollar y validar los distintos modelos. Además, para la simulación de muchos de ellos ha sido necesario el uso de ordenadores con una amplia capacidad en lo que respecta a memoria RAM.

Del mismo modo, aunque en dicho estudio la realización de una simulación está justificada de acuerdo a los dos problemas de investigación planteados en la tesis, el riesgo debido a que las condiciones planteadas en los distintos modelos sean poco realistas puede conllevar a limitaciones en lo que respecta a la generalización de los resultados (validez externa). Pues, aunque esta investigación se ha intentado ajustar en la medida posible a la realidad educativa, se podrían haber desarrollado más modelos vinculados con la misma.

Asimismo, se podrían haber simulado muchos otros modelos con muchas otras combinaciones, pero se ha optado por aquellos que se consideran que pueden tener un mayor interés e impacto en el estudio de la fiabilidad en la estimación de las puntuaciones de los estudiantes, así como en el análisis del comportamiento del ERM,

aunque en determinadas ocasiones no representen de la manera más fiel la realidad educativa.

Otra limitación es que en los modelos simulados no se manipula el error que se comete en cada una de las mediciones realizadas, sino, la varianza del error intrasujeto, que, en un mismo modelo, es siempre la misma en las diferentes ocasiones de medida. Pero en un futuro se pretende estudiar que es lo que ocurre cuando existe correlación entre los términos del error intrasujeto.

En esta investigación, se ha estudiado la fiabilidad en las estimaciones de las puntuaciones de los estudiantes, utilizando una estructura propia de los modelos jerárquicos lineales, pero en las evaluaciones educativas longitudinales, también es muy común, como ocurre en el EVAAS, el uso de modelos lineales mixtos con medidas repetidas. Por lo que sería muy interesante simular alguno de los modelos planteados en dicha investigación, con un diseño basado en modelos mixtos de medidas repetidas, ya que permitiría comparar los resultados obtenidos en las diferentes metodologías.

Además, en los modelos desarrollados no se han tenido en cuenta la introducción de ninguna variable predictora, a pesar de que las variables contextuales pueden afectar a las estimaciones de las puntuaciones de los estudiantes. Sería adecuado plantear una simulación en la cual se tuviera en cuenta ciertos predictores contextuales, como pueden ser el sexo, el nivel socioeconómico de los estudiantes, el clima del aula, etc., que son variables que afectan en el rendimiento de los estudiantes, y así, estudiar si la introducción de las mismas en el modelo afecta en la precisión con la que se estiman los niveles de logro.

En esta investigación, se estudia cómo el tamaño de las aulas (número de estudiantes por aula) afecta en la estimación de la pendiente de crecimiento de las mismas, pero sería interesante comprobar en un estudio futuro, el efecto de las estimaciones BLUP en el cálculo de los efectos aleatorios, cuando se tiene un gran número de aulas compuestas por pocos estudiantes, ya que, se consideran que dichas estimaciones producen el mencionado efecto de encogimiento, haciendo que las escuelas o aulas con tamaños muestrales pequeños, no se diferencien de la media global.

Finalmente, aunque el ERM no afecta a las puntuaciones finales de los modelos longitudinales con más de dos ocasiones de medida, en los que se siguen siempre a los mismos estudiantes, si es un artefacto muy presente en los modelos de ganancia y en los de cambio de cohorte, por lo que se requeriría de una investigación más en profundidad con el fin de desarrollar estrategias que permitan paliar dicho efecto. De tal manera que los resultados basados en estos modelos no se encuentren contaminados por este artefacto estadístico.

CHAPTER VII: CONCLUSIONS DISCUSSION, LIMITATIONS AND FUTURE PROSPECTS

The final chapter of this doctoral thesis aims to compile and discuss the main conclusions reached over the course of the research. This is done with reference to the two research problems raised and the conclusions and discussion of each specific objective are referred to in the context of each of the problems. This chapter also includes the limitations of the thesis and possible future lines of research which could be carried out as a continuation of this doctoral work.

VII.1. Problem 1. Study of reliability in estimations of student scores and of the behaviour of the RME when essential elements of the growth model (variance and correlation between the intercept and slope) are modified

Two specific objectives were proposed in the context of this first research problem, which aimed to find out how the percentage of dispersion fixed in the random coefficients of a longitudinal multilevel model, and the correlation between the parameters of intercept and slope, affect the reliability of estimations of student scores and the behaviour of the RME. The first of these sought to ensure an adequate recovery of the manipulated experimental conditions before the final results were generated, while the second focussed on studying the reliability of estimations of student scores and the behaviour of the RME by studying the correlation patterns generated between the levels of achievement over time.

The conclusions reached in each of these specific objectives are as follows:

VII.1.1. Specific Objective 1. To ensure adequate recovery of the manipulated experimental conditions

With regard to the recovery of the parameters defined when generating the sample (see section *VI. 1.1.1. Recovery of the sampling size*), all the models simulate a sample distribution made up of 10,000 students divided into 400 classes, where the minimum number of students per class is 20 and the maximum is 30, with the mean being 25 students. Therefore, the parameters defined previously, i.e. the total number of students (10,000), the minimum number of students per class (20 students), the maximum (30 students) and the mean number of students per class (25), were generated correctly, thereby ensuring adequate recovery of the sampling characteristics.

As far as the other parameters are concerned (see section *VI. 1.1.2. Recovery of variance and correlation parameters*), there is greater accuracy in the recovery thereof at student level compared to class level, which may be due to differences in the sampling size of the two levels, because while the class level is made up of 400 units, the student level consists of 10,000 units. Furthermore, the higher the fixed correlation between the intercept and slope, the greater the homogeneity between the values of the same variable. However, in spite of these small differences, and as evidenced by all the tests carried out, there is a high level of accuracy in the various models generated as regards the recovery of the various fixed conditions, and this guarantees that the final results are generated with a sufficient degree of validity and reliability.

In view of the results obtained in paragraph VI. 1.1.1. and VI. 1.1.2, we can conclude that all the experimental conditions manipulated in this first research problem were adequately recovered.

VII.1.2. Specific Objective 2. To study the reliability of estimates of student scores and the behaviour of the RME under different experimental conditions

The following section VII.1.2 presents the major conclusions reached on the basis of the results obtained in this second specific objective. The conclusions and the discussion thereof are structured with reference to each of the three situations explained in section *V. 3.3.3. (Model generation procedure)*.

VII.1.2.1. Conclusions and discussion of the models in Situation a

From the results obtained in ***situation a***, in which variability was only present in certain random coefficients of the longitudinal multilevel model, we can conclude the following:

On the basis of the models analysed in this situation, we can state that **the accuracy with which student scores are estimated is not affected by the level (class level or student level)** in which variance is found, provided that the dispersion present in the various random coefficients is always the same, i.e. if the data from two different models exhibit the same variance in the intra-subject error (for example, 95% of the total variance of the model) but in one of them there is only variability at the class level (5%) and in the other at the student level (5%), with the coefficients in which there is variance and the value thereof being the same in both models (Models 6 and 2, respectively), a very similar correlation matrix is obtained between student scores, even though the number of units at each level is different.

However, although the models in question provide very similar results, they represent very different interpretations. In the model where there is variance at the student level, the correlation values between the scores are due above all to the systematic characteristics thereof, whereas in the model where there is variance at the class level, the results are especially affected by the characteristics of the classes, but not of the students. **Therefore, similar results can be due to very different conditions (effects of the characteristics of the class or the students, or of both if there is variance at both levels)**, an aspect which no one had previously demonstrated.

Furthermore, the results obtained in the simulated models in *Situation a* show that when academic performance differs over time (there is variability in the slope, regardless of the level), even though the percentage of variability is very small (for example, 5%) in comparison with the total variance of the model (95%), **the correlation pattern observed is a cumulative pattern, i.e. as more measurements are made, the correlation values increase, these being the most reliable estimates**, as also demonstrated by Nesselroade et al. (1980) and later Rogosa and Willet (1985).

From the above we can deduce that **growth models do not always generate a simplex between measurement occasions**, as many authors (Humphreys, 1960; Joreskog, 1970, Joeskog & Sorborm, 1977 and Wert & Hilton, 1977) considered to be expectable.

Only in certain *Situation a* models, when initial achievement levels and performance over time differ between students (there is variance in the status and slope of a same level), do we observe that, within the first measurement occasion, the pattern tends to be a simplex. This tendency is maintained within other measurements, especially in the models of Sub-situation 8, where students start from different levels of average achievement and their growth curve is also different, with the variability found in these coefficients being very high compared with that of the intra-subject error. In these models (Model 8, 8.1 and 8.2), we observed that after the first measurement occasion, the greater the distance between two applications, the lower the correlation value between them. Thus, for example, the correlation between occasions two and three was higher than the correlation between occasions two and four and the latter was higher than the correlation between occasions two and five, and so on. Therefore, certain measurement occasions for these models resulted in a simplex pattern, to which many authors (Balzarini, Macchiavelli & Casanoves, 2005; Garson, 2013; Guttman, 1954; Humphreys, 1960, 1968; Kenny, 1979; Lunneborg & Lunnerborg, 1970; Pardo & Ruiz, 2013; Raudenbush. & Bryk, 1986; 2002; Ronald et al., 2010; Vallejo 1989; Vallejo & Fernandez, 1990) refer when discussing the expectable pattern between subjects' scores over time.

However, when a model is found where the percentage of variance of the intra-subject error is much higher (for example, 95%) than that of the variance of the intercept (for example, 2,5%) and slope (for example, 2,5%) (Model 7), the said simplex obtained in the previous models on the first measurement occasions, disappears and becomes a constant pattern with a certain upward tendency.

In short, the simplex pattern is expectable when the variance of the intra-subject error is very small (less than 35%) compared to the variance found in the other coefficients.

If the variance of the intra-subject error is very small (5%) compared with the variance of the slope (95%) (Model 3), we obtain reasonably reliable information about student scores with few measurement occasions, e.g. three. This makes it unnecessary to carry out a high number of measurements, because the limited additional information which they provide may not compensate for the cost involved. In contrast, as the variance of the intra-subject error increases with respect to the variance of the curve, the value of the correlations decreases, especially if the number of applications is low.

Thus, the reliability of the estimates in the models in which there is variability in the intra-subject error and in the curve depends on the amount of dispersion in each of these coefficients; the greater the variance in the intra-subject error, the less reliable the estimates of student scores.

With regard to the effect of the variance of the intercept, the results obtained enable us to conclude that, when there is no variability in the slope, but there is variability in the status, i.e. when the initial levels of student or class performance differ from each other (there is variability in the status) but this level of performance remains stable over time (there is no variability in the slope), the correlations between the first measurement occasion and the remaining measurements are higher than those obtained when the students or classes have the same initial performance but their level of achievement differs over the various applications (there is variability in the slope). However, in this latter situation where there is variance in the slope but not in the intercept, the correlations from the second measurement occasion onwards are higher than those obtained when the initial levels of student achievement are different but their growth rate is the same. **This is due to the fact**

that the random coefficients of the intercept are not cumulative, unlike the variance of the slope which is multiplied by time, and, therefore, as more measurements are carried out, the correlation values increase between one occasion and another.

In general, the correlation matrix generated when there is only variability in the intra-subject error and in the intercept is an extremely homogeneous matrix, where the reliability of estimates over time is very similar, as the growth curve has no effect on the various measurement occasions, and therefore a very similar score is always estimated. When the dispersion in the status is very high (for example, 95%) compared with the variance of the intra-subject error (for example, 5%) (Model 6 and Model 6.1), the correlations between the first measurement occasion and the rest of the occasions are very high, and these values remain constant over the entire correlation matrix. The opposite occurs when the percentage of variance of the intra-subject error is very high (for example, 95%), compared to that of the variance of the status (for example, 5%) (Model 5).

As far as the RME is concerned, the **constant pattern** between the first measurement occasion and the rest of the applications, which results in this effect remaining unchanged over time regardless of the number of measurements carried out (Nesselroade et al., 1980), was found in the models in this situation when all the classes or all the students started from similars levels of initial achievement there being all the variability in intra-subject error, when the students or classrooms performance is different throughout the applications (there was variability in the slope) but it remains constant at the first occasion of measurement (there is no variance in the intercept) or when there was variability only in the status and intrasubject error. The **simplex** between the first measurement occasion and the rest of the applications, which means that the RME becomes more pronounced as the number of measurements increases (Nesselroade et al., 1980), was found when there was dispersion in the intra-subject error and in all the random coefficients at student level or class level, above all where there was a significant difference between the amount of variance of the status and the amount of variance of the intra-subject error, with the latter coefficient being lower. Although the **increasing pattern**, which indicates that as more measurements are made, the value of the correlation between

student scores on the various measurement occasions increases, with the RME actually disappearing if the correlation between any of the measurements is perfect (Nesselroade et al., 1980), this upward tendency was not observed in any of the models of this situation between the first measurement occasion and the remaining applications, although a perfect correlation was found within the first measurement when there was no variability in the intra-subject error (Model 5.2).

The **most important ideas** which arise from *Situation a* are as follows:

1. The class or student level which exhibits variability does not affect the reliability with which student scores are estimated, provided that the amount of dispersion is equal in the same coefficients.
2. A cumulative pattern is observed between student scores in the different measurements when their performance or that of the classes varies over time.
3. The variance of the status, unlike that of the slope, does not have a cumulative effect; thus, when students or classes start from different levels of achievement, but these remain constant over time, an extremely homogeneous correlation matrix is obtained.
4. A simplex is not expectable every time we work with longitudinal data but only when the dispersion of the intra-subject error is small compared to the other coefficients.
5. When the percentage of variance of the slope is very high compared to that of the intra-subject error, fewer measurement occasions are required to obtain reliable estimates of student performance than when the situation is reversed.
6. With regard to the RME in *Situation a* models, an fundamentally constant and decreasing pattern was found between the first measurement occasion and the rest of the measurements. Thus, this phenomenon remains constant or increases over time except when there is no variance in the intra-subject error and all the variability is due to the intercept; in this case there is no RME, since the correlations between the first measurement occasion and the remaining applications are perfect.

On the basis of these conclusions, the following **suggestions** are made with a view to carrying out quality educational assessments.

a) In unrealistic conditions (with a very small residual, maximum 5%), we can consider three measurements to be sufficient to provide a good estimate of the actual performance of the subjects.
b) Generally, if the levels of student achievement differ over time (there is a large degree of variability in the slope), the more measurements made, the more reliable the estimates of student scores.
c) If student performance is constant over time, the reliability of the estimates of student scores will always be very similar, regardless of the number of tests performed. Therefore, there will be hardly any difference between applying three measurements or nine.

VII.1.2.2. Conclusions and discussion of Situation b models

On the basis of the results obtained in ***Situation b***, characterised by variability in the intra-subject error, variance in the intercept and slope of a same level and different correlation values between the intercept and slope of that level, we can draw the following conclusions.

The correlation values between the intercept and slope of one of the levels (class or students) affect the accuracy with which student scores are estimated, although not in such a pronounced way as the variability of the random coefficients.

When comparing models where the intensity of the relationship between the intercept and slope is different, but with the same direction, there are few differences in the results obtained. Thus, for example, the correlation matrix generated when there is no relationship between the intercept and slope, and that obtained when the relationship is low (0.2), is very similar. When there is a more pronounced tendency, in which students who start from a high level of achievement grow significantly over time, and the opposite is true for those who start from a low level of achievement (correlation of 0.9), scores are more reliable with fewer measurement occasions than

when a low correlation value is found between the intercept and slope (0.2). However, this difference, in general, is not very pronounced.

If we compare two models in which students who start from a high level of performance remain constant or grow little over time, and those who start from a low level of performance increase this level significantly over the course of the measurements, with this tendency being not very pronounced in one case (with a correlation value between the intercept and slope of -0.2) and more pronounced in the other (where we find a correlation value of -0.9), the matrices generated in both models tend to differ more from each other than when we compare two models where the relationship between the intercept and slope is positive. Of course, the models which differ most from each other are those where there is a highly intense relationship between status and slope, but where one of them is positive and the other is negative.

When we study models where there are different correlation values between the intercept and slope, the differences in the reliability of the estimates of student scores between models are more pronounced when the number of measurements is small (three or four), because when the number is high we still observe the cumulative effect of the slope. Generally, therefore, **with a larger number of measurement occasions (more than five) the effect of the correlation value between the intercept and slope is not very pronounced, due to the fact that we obtain similar levels of accuracy both with five measurements and with nine.**

In the case of models with different correlation values between the intercept and slope, we observed that more reliable estimates of student scores are obtained with a lower number of measurement occasions in those models where the initial performance of the students or classes is closely related to the one they achieve over time. It is the students or classes with high initial achievement levels which grow the most over the different measurement occasions, and, in contrast, the classes or students with a low initial performance which increase these levels least i.e. when there is a high positive correlation between the intercept and slope (0,9).

Conversely, the models which provide less reliable performance estimates with a greater number of measurements are the compensatory models, in which the performance of students or classes with high initial achievement levels remains constant or does not grow much over time, while in the majority of cases of students or classes with low initial levels of achievement, their performance grows significantly over the various measurement occasions, above all if there is a strong association between the initial performance and the one they achieve over time. Therefore, when there is a low negative correlation between the intercept and the slope (-0.9), estimates are less reliable, especially if the number of measurements is low.

Both the correlation pattern obtained between the first measurement occasion and the rest of the applications, and those between the other measurements are the same as those obtained in the models which exhibit the same random coefficients but where there was no relationship between the intercept and slope. This pattern only changed when there is a high positive correlation between the intercept and the slope (0.9), because under this condition most models exhibit a pattern of correlation between the various measurement occasions which tends to increase. **In those models where an increasing pattern is obtained between the first measurement occasion and the other measurements, this increasing pattern is maintained or becomes constant within the remaining applications**. **However, when there is a low positive correlation value between the intercept and slope (0.2) or low and high negative (-0.2 and-0.9), then we do observe an autoregressive or simplex pattern between the first measurement occasion and the rest of the occasions, and this extends to other measurements beyond the first application.**

The only model where an increasing pattern was not found between the first measurement occasion and the rest of the occasions, or between the correlations of occasions two and three and the rest of the measurements, is when there is a correlation of 0.9 between the intercept and slope, and the variability of the intra-subject error is very small (5%) compared with that of the intercept (47.5%) and the slope (47.5%) (Model 12.1). In this case, the correlation pattern within the first three measurement occasions tends to be a simplex.

With regard to the RME, the **constant pattern** was found in the models in this situation when the level of variance of the intra-subject error was very high (95%) compared with the dispersion of the intercept (2.5%) and, furthermore, the correlation between the intercept and slope was 0.2, -0.2 or-0.9 (Model 10, Model 10.2 and Model 10.3). The **simplex** was obtained when there was variability in the intra-subject error, in the intercept and in the slope of a same level, but where the variance of the intra-subject error was either not much greater than the variance of the status (Sub-situations 11 and 13) or was lower (Sub-situations 12 and 14) and, furthermore, the relationship between the intercept and slope was low positive (0.2), low negative (-0.2) or high negative (-0.9). Where there was a high positive correlation, a simplex was only found where the percentage of variance of the status was very high compared to the percentage of variance of the intra-subject error (Model 14.1). Lastly, the **increasing pattern** was generated when the correlation between the intercept and slope was very high and positive (0.9), unless the variance of the intra-subject error was very small compared with the variance of the status (Model 12.3).

The most important **main ideas** obtained from the models generated in this situation are as follows:

1. The relationship between the starting point and the growth curve of the student or class level affects the reliability with which student scores are estimated, but not in such a pronounced way as the dispersion of the random coefficients.
2. When the strength of association between initial performance and the growth curve differs between models, but the direction of the relationship is the same, the accuracy with which student scores are estimated is very similar in all models.
3. When we compare models where the direction between the starting point and the growth curve is different, we obtain more accurate estimates with fewer measurement occasions in the models where the initial performance is related to the performance achieved over time (positive direction).
4. The effect of the relationship between the starting point and the growth curve on the accuracy with which student scores are estimated over time is

more pronounced when the number of measurements carried out is small (three or four applications).

5. The models which provide more information with fewer measurement occasions are those where the intensity of the relationship between initial performance and growth is high and positive; however, the least reliable are those where the correlation between status and slope is also high but negative (compensatory model).
6. When there is a very pronounced tendency where students with a high initial level of performance progress significantly over the applications, and the reverse is true for those with a low initial performance, we do not observe a simplex pattern but rather a predisposition towards an increasing pattern within the various measurements. This means that the RME could even disappear if the correlation between the first measurement opportunity and any of the others were perfect.

On the basis of the conclusions obtained in S*ituation b*, **the following considerations** should be taken into account by experts and those interested in quality education assessments:

a) In assessments where there is a pronounced tendency for the initial performance of the students or schools not to be closely related to the performance achieved over time, and where there is a balanced variability between the various random coefficients (e.g. Model 13.3), at least five or six measurements are required in order to obtain accurate estimates of student performance (correlations above 0.90). However, in this same situation when there is a positive relationship between status and slope, student scores can be estimated for subsequent courses with only three or four measurements, as we generally obtain correlations between student scores which are greater than 0.90.
b) Therefore, the compensatory classes or schools which seek to even out the levels of achievement of their students, i.e. so that those who start from a previously low level of performance will grow significantly over time, and those who start from a previously high level of performance remain

constant between applications, require a greater number of measurements to estimate student performance with a certain degree of reliability than those schools where the initial performance of the students is closely related to the performance acquired over time.

VII.1.2.3. Conclusions and discussion of Situation C models

On the basis of the results generated in ***Situation C*** models, where there was variability in all the random coefficients, and different correlation values between the intercept and slope of student level and class level, we can draw the following conclusions.

From the correlation matrix obtained in Model 15, which was generated on the basis of the values of variance found in a number of multilevel longitudinal research studies (Castro et al., 2009; Navarro, 2013; Raudenbush & Bryk, 2002; Zvoch & Stevens, 2003; Zvoch & Stevens, 2008), we can conclude that, **in order to estimate student performance with a certain degree of accuracy (correlations greater than 0.90), at least three measurements are required, and this is the minimum number which permits a growth model to be generated. Extremely reliable estimates of student performance are obtained with four applications, as correlations stood at around 0.95**.

In Model 15, little difference was found between the information provided by five measurements (correlation of 0.97) and that obtained with nine (correlation of 0.99); in this model, therefore, as is the case in many others, **carrying out more than four or five applications is not cost-effective, because the extra information provided does not compensate for the cost involved.**

In these *Situation C* models, where both students and classes start from different levels of achievement, and their performance over time also differs, when two models exhibit the same variability, the same random coefficients and the same correlation values between the intercept and slope, but in reverse order (e.g. in one model the correlation between the intercept and slope of the class level is 0.2 and that of the student level is 0.9, and in another model the opposite is true, i.e. the correlation between status and slope of the student level is 0.2 and of the class level it is 0.9), we observe that **the correlation values between student scores in both models are**

generally very similar. However, in some specific cases we may find that correlations are slightly higher in the model in which the highest correlation is at student level, and this may be due to the fact that there were more units of analysis at this level (10,000 students) than at level three (400 classes).

When we compare the results obtained in a model in which there is only a correlation between the intercept and slope of a single level, with a model where, in addition to having the same correlation value at that same level (class or student), there is a correlation value between the intercept and slope of the other level, **the reliability of the estimates in both models is generally very similar**. This may be due to the fact that, as indicated above, the correlation value between the intercept and slope does not have a very pronounced effect on the reliability with which student scores are estimated over time.

In the same way, when we compare the results of models in which there is the same dispersion in the random coefficients, but different correlation values between the intercept and slope of student level and class level, if the direction of the fixed correlation values between the intercept and slope is the same at both levels in the different models, **the results provided by these models are very similar, even if the intensity of the fixed correlation values between the intercept and slope is very different.** Therefore, there is not much difference between the correlation matrix obtained in a model where the correlation between the intercept and slope of the class level and student level is 0.2 (Combination 1), and that obtained when these correlation values stand at 0.9 (Combination 6). Although this difference in correlation matrix values is more pronounced if the said values are negative (Combination 11 and Combination 16, respectively).

Of all the combinations in *Situation C*, it is **Combination 6 which offers the most reliable estimates with a lower number of measurement occasions**, because the initial performance of the classes and students was closely related to the levels of achievement acquired over time (the correlation between the intercept and slope of student level and class level was 0.9).

When these correlation values between status and slope are found at class and student level (0.9), and there is a balanced variability between the different random coefficients (Sub-situation 17), correlations greater than 0.90 are obtained between

student scores on the various measurement occasions with just three or four measurements. In contrast, **the combination that exhibits the lowest degree of accuracy when estimating levels of student achievement is Sub-situation 16, as the correlation between the intercept and slope of class level and student level was -0.9**. In the model where the variability of the random coefficients is based on educational reality and reflects a correlation of -0.9 between the intercept and slope at class and student level (Combination 16 of Sub-situation 17), between 6 and 7 measurements are required to obtain reliable correlations (greater than 0.90).

As far as the RME is concerned, in this situation we found a **constant pattern** when the percentage of variance of the intra-subject error accounted for 95% of the total variance of the model and the remaining 5% was distributed equally between the other random coefficients (Sub-situation 18). With these percentages of variance, the RME remains constant in all combinations, with the exception of Combination 6, in which there was a high positive correlation (0.9) between the intercept and slope of student level and class level. In this case, we observed a certain tendency towards an increasing pattern. With regard to the **simplex**, this was found in Model 15, which was generated from the results obtained in a number of educational research studies (Castro et al., 2009; Navarro, 2013; Raudenbush & Bryk, 2002; Zvoch & Stevens, 2003; Zvoch & Stevens, 2008), in the model generated from Raudenbush and Bryk's original model (2002) (Model 16), in all combinations of Model 17, in Model 18 (except in Combination 6) and in Model 19. The **increasing pattern** appeared in Combination 6 of Sub-situation 18, where the correlation between the intercept and slope of the class level and student level was 0.9.

The **key ideas** arising from this situation are summarised below:

1. The level (student or class) where a higher correlation value is found between the intercept and slope does not significantly affect the reliability with which student scores are estimated.
2. There is hardly any difference in the accuracy with which student scores are estimated when there is only a relationship between initial performance and that obtained over time at one of the levels (class or student) and when it exists at both (class and student).
3. Where there are different correlation values between the intercept and slope of class level and student level in two models, but the direction is the same, the reliability with which student scores are estimated is very similar in both models. Therefore, it seems that the intensity of the relationship between the intercept and slope does not have a significant effect on the accuracy with which student scores are estimated.
4. When there is a relationship between the intercept and slope of the class level and student level, the models which provide most information with a smaller number of measurement occasions are those where both the students' starting performance and that of the classes is closely related to their growth curve. In these models, student scores can be estimated with a high level of accuracy with just three or four applications. In contrast, when there is a high negative correlation between the intercept and slope of class level and student level (-0.9), at least six or seven measurements are required to obtain accurate estimates of student performance.
5. Generally, in the various combinations of the *Situation C* models, the correlations between the first measurement occasion and the rest decrease. Therefore, the RME increases as more applications are carried out.

From the results obtained in *Situation C*, but, above all, from the results of Model 15, which was a continuation of other research studies, we offer the following **recommendations**:

a) On the basis of the results of Model 15, student performance can be accurately estimated with three applications, as correlations are greater than 0.90, but with four measurement occasions correlations are greater than 0.95.
b) Sometimes it is not cost-effective to carry out a very high number of measurements, e.g. more than four, because they do not provide much additional information. More than four applications are especially necessary when working with data where initial performance bears little relation to growth. In these cases, especially where there is a high-intensity correlation between the intercept and slope of class level and student level, at least six or seven measurements are required in order to be able to estimate student performance accurately (correlations greater than 0.90).

VII.1.2.4. General conclusions drawn from Situation A, B and C models

The **general conclusions** obtained from all the simulated models in *Situations a, b* and *c,* are as follows:

1. **The factor which most affects the reliability of estimates of student scores is the difference in dispersion between intra-subject error and slope**. The greater the variance of the slope compared to the variance of the intra-subject error, the more information on student scores is obtained with fewer measurement occasions, regardless of whether the dispersion of the slope is at student level or class level.
2. **The variance of the intra-subject error compared to the dispersion of the other random coefficients plays a very important role in the accuracy with which levels of student achievement are estimated over time**. When all the variance is accumulated in the intra-subject error, considering that all the students who attend the same class, and the classes themselves, start from the same average level of performance and their growth is constant over time, the correlations between student scores on the different measurement

occasions are nil, because there is no variability either in the status or the intercept.

3. **When there is no dispersion in the intra-subject error, we obtain a perfect correlation pattern between student scores**, whether all the variability is in the intercept, i.e. whether the classes or students start from different initial performance levels or whether students or classes grow at different rates (there is variance in the slope). In this latter situation, a perfect correlation pattern is generated, except between the correlations of the first measurement occasion and the rest of the measurements, since the initial level of performance is the same for all students and all classes.
4. If the variance of the intra-subject error is lower than that of the other random coefficients, more information is obtained on student performance than when the situation is the reverse, especially when the number of measurements is not high. Thus, **if the variance of the intra-subject error is high (95%) compared to the value of the other random coefficients, an accurate estimate of student scores will not be obtained even with nine measurements (correlations below 0.90). However, if the variance of the error is small (5%), we can obtain correlations between student scores greater than 0.90 with just three measurements.**
5. **The correlation values between the intercept and slope do not have a very pronounced effect on the reliability with which student scores are estimated over time**, although the direction and the intensity of the relationship between status and slope provide very different interpretations between the relationship of the initial performance and that which is achieved over time.
6. Another of the major contributions made by this thesis, obtained from the results generated in the various simulations, is that, on the basis of the correlation patterns found, **it is a mistake to consider that, when working with repeated measurements, we can expect a simplex or autoregressive pattern** as many authors indicate (Garson, 2013; Pardo & Ruiz, 2013; Raudenbush & Bryk, 2002; Ronald et al., 2010, etc.). This pattern appears in a pronounced way within several measurement occasions, when both the initial performance and the one achieved over time are different (there is dispersion

in intercept and slope of class level, student level or both) and, in addition, there is much greater variability in these coefficients compared to the variance of the intra-subject error. As the percentages of variance of status and slope approach and match that of the variance of the intra-subject error, this pattern diminishes, and it disappears when the variance of the intra-subject error is much greater than the total dispersion of the sum of the variance of intercept and slope, unless there is a negative correlation value between the intercept and slope.

When students and/or classes with a high initial level of achievement remain constant over time or do not grow much, but the performance of students and/or classes with low initial levels of achievement does increase significantly, i.e. there is a compensatory relationship between the intercept and slope, we can also find a simplex pattern between certain measurement occasions.

7. Although in certain models **it is possible to find a simplex within some measurement occasions, in almost all the simulated situations (including the models that exhibit an autoregressive pattern within any measurement occasion)**, **we can observe an increase in the correlation values (increasing pattern) as more measurements are added.** Thus, for example, although we may find that, in some models, the correlation between the first and second measurements is greater than the correlation between the first occasion and the third, and that this correlation, in turn, is greater than that of the first occasion and the fourth, in all the models where the performance of the students or classes differs over time, the correlation between occasion one and two is lower than the correlation between occasion two and three, and this in turn is lower than the correlation between occasion three and four, and so on, in spite of the fact the time lapse between each measurement is the same, as there is a cumulative effect in the growth curve.
8. **The fact that the pattern between student scores over time increases as more applications are carried out indicates that the characteristics of the students and/or classes have an effect on performance**. If the pattern were a simplex it would mean that the characteristics of the students and/or classes had no effect on levels of achievement. This would represent a situation

similar to the one obtained when students change class or school continuously between measurements and their results were not related to each other. Thus, we can expect the autoregressive matrix or simplex as more applications are added, provided that the classes or schools do not intervene in the learning process and that this absence of intervention is not offset by the student's own academic progress.

9. **The most reliable model with the lowest number of measurement occasions (3 applications) is the one where there is little variability in the intra-subject error compared with that of the growth curve, and where there is a high positive relationship between the intercept and slope of both levels.**
10. Raudenbush and Bryk (2002) consider that two points in time are sufficient in order to study progress between one application and another, and that, with three, it is possible to estimate a growth curve. However, when we seek to ensure the reliability of estimates of student scores, the number of measurement occasions required depends, above all, on the difference between the variability of the intra-subject error and the growth curve, because, when the percentage of variance of the intra-subject error is very high, the effect of the variance of the slope is observed in the long term. However, **when working with values inherent to educational reality (Model 15) where there is no significant difference between the two percentages of variance, we usually obtain fairly reliable estimates with just four measurements**. As demonstrated in Model 15, the correlations between student scores over four measurement occasions are 0.95. Therefore, assessments such as those carried out by the NWEA, which are based on five measurement occasions, seem to be very reliable and make it possible to estimate student performance accurately.
11. With regard to the RME between the first measurement occasion and the rest of the applications, in **the majority of simulated models this phenomenon remains constant or increases over the various measurement occasions**, and only tends to disappear (increasing pattern) when there is a high positive relationship between the intercept and slope (0.9).

As regards possible **general recommendations** based on the conclusions reached in this first research problem, and in order to improve the quality of educational assessments, the following are suggested:

a) When, in an educational assessment, little variability in the intra-subject error is observed compared with the variance of the growth curve, and there is also a positive relationship between initial performance and the performance achieved over time at both levels, then we are dealing with a model where three measurement occasions will be sufficient to permit an extremely reliable estimate of student performance.

b) In general, we recommend at least four applications in the educational assessments which monitor the same students in order to be able to estimate their performance as accurately as possible. We are aware that this is not always possible, among other reasons, due to the cost involved. In this case, it is advisable to bear in mind the limitations that this may pose, especially if the assessments are of a high-stakes nature, where schools which do not achieve previously established minimum levels of performance are penalized. For this reason, and for many others related to the complexity involved in measuring levels of achievement, we would encourage assessments of a diagnostic (low-stakes) nature aimed at introducing improvements in education systems in accordance with their limitations and needs.

c) Educational assessments must be understood as a tool which helps to reinforce the strengths of educational systems and acts as an aid to overcoming difficulties, and always for the purpose of improvement. They should never be conceived as a means of punishing schools for their limitations. A school is not a company where all that matters is end performance. At school the important thing is where a student starts from and how far he or she gets within a specific context. The fundamental thing is how students and schools change over time. Not all students reach the same level because neither do they all start from the same point. In this sense, it is longitudinal evaluations, which monitor the same students over time, which provide the most information on progress, both of both schools

and students, and, therefore, will permit the implementation of programs aimed at effectively improving educational quality.

d) Educational assessments which monitor the same subjects over more than two measurement occasions should not concern themselves with the RME in their final results as this phenomenon only occurs between the first measurement occasion and the rest of the occasions. However, for assessments based on progress models (two occasions) or on models where new subjects are selected over time, the behaviour of the RME depends on the correlation pattern found between the various applications. If the pattern is constant, the RME will remain unchanged over the various measurement occasions. In the case of a simplex, the RME will increase over time. And in the case of an increasing pattern, the RME may even disappear if a perfect correlation between the first measurement and another measurement occasion is obtained.

VII.1.2.4. Conclusions of the study of the performance of students grouped in different quintiles

Before ending this section on conclusions and discussion relating to the first research problem, it is worth indicating the most important contributions made when seeking to address the final question, posed as a postscript to this research problem: **Which classes most foster the performance of the students grouped into the various quintiles?**

To address this question, the four models in Sub-situation 14 were used, where the intra-subject error was fixed at 5%, there was no variability in the student growth curve and the variance of the intercept of that same level was fixed at 15%, that of the class level at 55% and its growth curve at 25%. As for the correlation between the intercept and slope of the class level, this was fixed at 0.2 in Model 14, 0.9 in Model 14.1, -0.2 in Model 14.2 and -0.9 in Model 14.3.

On the basis of the results obtained in the classes in these models grouped in the five quintiles, if it were possible and/or necessary to select a class situated within the **first quintile**, the ideal approach would be to select classes in which initial performance were not closely related to performance achieved over time, i.e. compensatory classes, where those with a low initial level of performance progress

notably over time and the reverse is true for classes with a high initial level of achievement **(Model 14.3 classes)**. The choice of these classes in the first quintile is because the results show that in Model 14.3, regardless of the type of class (low, medium or high slope) and the quintile in which students are situated, all of them achieve a growth rate higher than that obtained in the rest of the models. However, of the models in Sub-situation 14, it is this model 14.3 which is the one which requires most measurements in order to obtain reliable estimates of student achievement. Specifically, four to five measurements resulted in a correlation value of 0.94, while five or six measures increased this value to 0.97.

Of the classes belonging to the **second quintile**, when the slope residual (value-added) is low, better results are also obtained in the classes where initial performance is not closely related to performance achieved over time **(Model 14.3)**, but when the slope residual is medium or high, the results are better in those classes where initial performance is related to performance achieved over time, though this trend is not very pronounced **(Model 14).** In this model, five or six measurements produced a correlation value greater than 0.97.

As for the classes in the **third quintile**, if they have a low slope residual, students exhibit a higher rate of growth if they attend classes in which there is a slight tendency for the initial level of achievement not to be closely related to the level achieved over time **(Model 14.2)**. In this model, a correlation value of 0.95 was obtained with three or four measurements. However, when the slope residual is medium or high, progress is greater if the students attend classes in which there is a slight tendency for the initial level of achievement to be related to the level achieved over the course of the measurements **(Model 14)**.

Within the classrooms located in the **fourth quintile**, in the classrooms with a low slope residual, students increase their performance more when they attend classes which exhibit a slight tendency for initial levels of achievement not to be closely related to their growth curve **(Model 14.2)**. However, when the residual of the growth curve is medium or high, it is better for them to attend classes where the initial performance is related to performance achieved over time, without this tendency being very pronounced **(Model 14.1)** or classrooms where there is a slight tendency in which the initial achievement levels are not closely related to those

obtained throughout the rest of applications **(model 14.2).** If the slope residual is high, students obtain a higher gain when they go to classrooms that have a relationship between the intercept and the slope is positive and low intensity **(model 14).**

Finally, with regard to the classes belonging to the **fifth quintile**, the students perform better when the slope residual is low or medium in classes where there is a slight tendency for initial levels of achievement not to bear much relation to those one achieved over time **(Model 14.2)**. However, when the slope residual is high, there is little difference between attending classes where the starting performance is related to the performance acquired over time, without this tendency being very pronounced **(Model 14)**, and attending classes exhibiting the same tendency but in a more pronounced way **(Model 14.1)**. However, the two models differ in the number of measurements required to obtain accurate estimates of student performance, as we must remember that Model 14.1 requires less measurements than Model 14. In this latter model, correlations of 0.97 were obtained with three or four measurements. This same value was obtained in Model 14.1 with two to three measurement occasions.

The main **key ideas** arising from these conclusions are as follows:

1. Although, in general, **Model 14**, with its low positive correlation (0.2) between the intercept and slope of the class level, is the model least detrimental to the performance of students, since always within some of the classrooms located in the different quintiles there are students who obtain a positive gain between the first occasion of measurement and the last. The classrooms of quintile five of **Model 14.1** and **Model 14.2,** together with the classrooms of quintile one of **Model 14.3**, are most promote the growth of students from all quintiles.
2. In contrast, it is in the model where initial performance is related to performance achieved over time, without this tendency being very pronounced **(Model 14.1)**, where the classes of the different quintiles generally exhibit the worst performance. It is the students attending classes in the **first quintile of this model** who exhibit the worst overall rate of performance between the first and last measurement occasion, as all the

students situated in the various quintiles of these classes experience a fall in their level of achievement between the first and last application. Although the classrooms located in the first quintile where the correlation between the intercept and the slope is low of negative directionality (**Model 14.2**) and in the classrooms of the fifth quintile of the model where the correlation between the intercept and the slope of the student level it is also negative, but of high intensity (**Model 14.3**) it has been observed that students from all quintiles have a significant loss in their levels of achievement over time.

3. Although, within the same class, there is not much difference between the growth rate of students grouped in different quintiles, there is a big difference in the growth rate of students who attend classes where the relationship between the intercept and slope differs in direction and intensity. **This correlation value has a significant impact on the student growth rate between the first measurement occasion and the rest of the occasions.** However, this factor did not much affect the number of measurements required in order to make reliable estimates of student academic performance.

These ideas lead us to make the following **recommendations**:

a) Due to the importance of the intensity of relationship and the direction between the initial performance of the classes and their growth curve for the rate of student progress, it would, if possible, be worth studying this aspect in the different classes in the schools, so that students can attend those classes which most foster their academic performance.
b) We recommend reinforcing compensatory classes, since, in addition to leveling the students' performance, making the students whose initial performance is low, grows a lot over time and those who start from high levels of achievement keep constant, are the classrooms where students increase their growth rate and less loss experience between the first measurement occasion and the last.

c) We recommend studying and working with those classrooms located in the different quintiles, where the growth rate of students belonging to the different quintiles decreases as more applications are made.

VII.2. Problem 2. Effect of the sampling size of the classes on the estimation of their growth curve

In this second research problem, which was aimed at finding out how the size of the class affects the estimation of its growth curve, two specific objectives were established. The first of these was to ensure an adequate recovery of the manipulated experimental characteristics before the final results were generated and the second focussed on studying the effect of the size of a class on the accuracy with which its growth curve is estimated, under three different circumstances: with small, medium and large classes.

The conclusions reached in each of these specific objectives are presented below:

VII.2.1. Specific objective 1. To ensure adequate recovery of the manipulated experimental conditions

In view of the results obtained in *Section VI. 2.1.* with reference to the first specific objective of this second research problem, we can conclude that the experimental conditions established in each of the situations (*Situation A, B and C*) or models proposed (Model 15.a, Model 15.b and Model 15.c) were generated correctly, as in all situations a percentage of classes has been simulated which was very close to one defined previously, with the minimum and maximum number specified in each type of class (small, medium and large). Only in the large classes, where the intention was to generate a maximum of 30 students per class, was a maximum of 31 students generated in *Situations a* and *c,* and a maximum of 29 in *Situation B*, which may be due to the fact that sampling distributions are generated at random.

This does not present a problem when it comes to addressing the second specific research objective, because in large classes the important factor is not the maximum size generated, but rather that the minimum is greater than 19 students, as many authors (Goldstein, 1997; Bryk & Raudenbush, 2002; Raudenbush & Bryk, 1986) consider that 20 is the minimum number of units required in order to carry out

a multilevel model. Furthermore, educational legislation (LOMCE, 2016) establishes the maximum number of students in compulsory secondary education as 30 per class, with the possibility of raising this number by up to 10% to meet immediate schooling needs.

In view of what has been stated above in this section, we can conclude that all the sampling conditions manipulated in order to address this second research problem were adequately recovered

VII.2.2. Specific Objective 2. To study of the effect of the sampling size of the classes on their growth curve

Based on the results obtained (*Section VI. 3.1.*) in each of the situations generated, the following conclusions can be drawn:

In ***Situation a,*** where 74.5% of the classes were small (fewer than 11 students per class), 12.2% were medium-sized (fewer than 20 students per class) and 13.3% were large (more than 19 students), **we observed that in the case of the small classes, there was a significant difference between the percentage of classes which did not differ from the mean (86.3%) and those that did (13.7%). This difference between the classes which do not differ from the mean and those that do was not so pronounced in the medium-sized classes (a difference of 31.4 percentage points), and even less pronounced in large classes (a difference of 4.2% percentage points).**

In ***Situation b*****,** where the number of small classes fell to 12.8% of the sample and that of medium-sized classes increased to 74.8%, **the same tendency was observed whereby, as the number of students per class increases, fewer differences are observed between the percentage of classes that differ from the mean and the percentage that does not.**

In ***Situation c*****,** where 11.3% of the classes were medium-sized, 12.7% were small and 76% were large it, what stands out is that, **although within each type of class (small, medium-sized and large) the difference between the classes that do not differ from the mean and those that do is not lower than in the other situations (*a* and *b*), we did observe that this is the situation with the highest percentage**

of total classes which differ from the mean; specifically 35%, whereas in *Situation a* this percentage stood at 20.7% and at 30% in *Situation c.*

Thus, the **main conclusions** to be drawn from this second specific objective are as follows:

1. The bigger the classes, the higher the percentage of the classes that differ from the mean.
2. As class size increases, there is a smaller discrepancy between the percentage of classes that differ from the mean and those that do not, and this difference is particularly small when there are more than 19 students per class. Therefore, the recommendation made by Bryk and Raudenbush (2002), Goldstein, (1997), Raudenbush and Bryk, (1986) and Snijders and Bosker (2012) that sample classes should contain at least 20 students to prevent bias in the estimates of their growth curve appears justified.
3. The percentage of classes of each type (small, medium-sized or large) does not affect the fact that there is a smaller discrepancy between the percentage of classes which differ from the mean.
4. However, there is a substantial leap between the total number of classes that differed from the mean when around 75% of the classes were small (*Situation* a) and when this was the percentage of large classes (*Situation c*). About 15% more classes differed from the mean in this latter situation. This difference is not so pronounced when we compare *Situations b* and *c,* with 5% more classes differing from the mean in *Situation c*.

On the basis of what has been discussed in this second specific objective, the following **recommendations** are put forward regarding the importance of sampling size when carrying out educational assessments:

a) We recommend carrying out assessments of classes with more than 19 students, as this is when there is a higher percentage of classes which differ from the mean and when there is least discrepancy between the percentage of classes that differ from the mean and the percentage that does not. If this is not possible, it would be advisable to take into account the impact of working with this type of classes when it comes to estimating their growth curve.
b) We do not recommend taking decisions, especially those of a high-stakes nature, based on the growth rate of classes containing a low number of students (fewer than 20), as in many cases it will not be possible to know their actual performance, as they will tend not to differ from the mean due to their estimation error.
c) We recommend a great deal of caution when drawing up school rankings based on estimates of the growth curve of this type of small and medium-sized class, since estimates of these growth curves contain so much error that the majority do not differ from the mean.

VII.3. Limitations and future prospects

First of all, it must be said that all research, even the most thorough, has limitations.

Specifically, in this thesis, a simulation of this magnitude, with three levels of nesting, 1,000 replicas and a wide range of combinations, requires large amount of computing time to prepare and validate the various models. Furthermore, simulating many of these models requires the use of computers with high-capacity RAM memory.

Likewise, although a simulation is justified in view of the two research problems posed in the thesis, the risk that the conditions established in the various models are unrealistic may give rise to limitations when it comes to extrapolating the

results (external validation). Thus, although we have attempted to adapt this research as far as possible to educational reality, we could have drawn up more models associated with that reality.

Furthermore, we could have simulated many other models with many other combinations but opted for those which we considered to be of greater interest and to have a greater impact on the study of reliability in the estimation of student scores and on the analysis of the behaviour of the RME, even if, on certain occasions, this does not reflect educational reality in the most faithful way.

Another limitation is that the error committed in each of the measurements applied in the simulated models is not manipulated, but rather the variance of the intra-subject error, which, for any given model, is always the same on every measurement occasion *is*. However, in the future we intend to study what happens when a correlation exists between the terms of the intra-subject error.

In this research, we have studied the reliability of estimates of student scores, using a structure typical of hierarchical linear models, but in longitudinal educational assessments, the use of linear mixed models with repeated measurements is also very common, as is the case of the EVAAS. It would therefore be very interesting to simulate some of the models proposed in the said research with a design based on mixed models with repeated measurements, as this would permit a comparison of the results obtained by the different methodologies.

Furthermore, the models generated during the course of this research did not take into account the introduction of any predictor variable, despite the fact that contextual variables can affect the estimates of student scores. It would be useful to consider a simulation which takes into account certain contextual predictors, such as sex, the socioeconomic status of the students, the atmosphere in the class, etc., as these variables affect student performance, and study whether their inclusion in the model affects the accuracy with which levels of student achievement are estimated.

In this research, we study how class size (the number of students per class) affects estimates of their growth curve, but it would be interesting to verify, in a future study, the effect of BLUP estimates in the calculation of random effects when working with a large number of classes made up of few students, because these estimates are

considered to produce a shrinking effect, whereby schools or classes with small sampling sizes do not differ from the global mean.

Finally, although the RME does not affect the final scores in longitudinal models which monitor the same students over more than two measurement occasions, it *is* an artefact with a strong presence in progress models and those based on change of cohort, and therefore requires more in-depth study in order to develop strategies which alleviate this effect, in such a way that the results based on these models are not contaminated by this statistical artefact.

REFERENCIAS BIBLIOGRÁFICAS

Abad, F., Olea, J., Ponsoda, V., & García, C. (2011). *Medición en Ciencias Sociales y de la Salud*. Madrid: Síntesis.

Achtenhagen, F. (2005). Competence and their development: cognition, motivation, meta-cognition. En W. Nijhof & L. Nieuwenhuis (Eds.), *The Learning Potential of the Workplace* (pp. 331-345). Twente, University of Twente.

Achtenhagen, F., & Grubb, N.W. (2001). Vocational and Occupational Education: pedagogical complexity, institutional diversity. En V. Richardson (Ed.), *Handbook of Research on Teaching* (pp. 25-68). Washington: American Educational Research Association.

Aitkin, M., & Longford, N. (1986). Statistical modeling issues in school effectiveness studies. *Journal of the Royal Statistical Society, Series A, 149*, 1-43.

American Educational Research Association, Américan Psicological Association & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* Washington: American Educational Research Association, Américan Psicological Association & National Council on Measurement in Education.

Anderson, J. A. (2005). *Accountability in education*. Paris: International Institute for Educational Planning.

Angoff, W. H. (1971). Scales, norms and equivalent scores (2nd ed.). En R. L. Thorndike, *Educational Measurement* (pp. 508-600). Washington, DC: American Council on Education.

Armein, A. (2008). Methodological Concerns About the Education Value-Added Assessment System. *Educational Measurement, 37* (2), 65-75.

Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R., Rothstein, R. (2010). *Problems with the use of student test scores to evaluate teachers*. Washington, D.C.: Economic Policy Institute. Recuperado de epi.3cdn.net/b9667271ee6c154195_t9m6iij8k.pdf

Ballou, D. (2002). Sizing up test scores. Education next. *Education next Summer*, 10-15. Recuperado de www.cgp.upenn.edu/pdf/10.pdf

Ballou, D. (2009). Test Scaling and Value-Added Measurement. *Education Finance & Policy, 4* (4), 351-383.

Ballou, D., Sanders, W. & Wright, P. (2004). Controlling for student background in value-added assessment of Teacher. *Journal of Educational and Behavioral Statistics*, *29* (1), 37–65.

Baltes, P. B., & Nesselroade, J. R. (1979). History and rationale of longitudinal research. En J. R. Nesselroade & P. B. Baltes (eds.), *Longitudinal research in the study of behavior and development* (pp. 1–39). New York: Academic Press.

Baltes, P. B., Reese, H. W., & Nesselroade, J. R. (1977). *Life-span developmental psychology: Introduction to research methods*. Monterey, CA: Brooks/Cole

Balzarini M, Macchiavelli R., & Casanoves, F. (2005). *Aplicaciones de modelos mixtos en agricultura y forestaría*. Centro agronómico tropical de investigación y enseñanza CATIE. Costa Rica.

Barton, P., & Coley, R. (1998). *Growth in school: Achievement gains from the fourth to the eighthgrade*. Princeton, NJ: Educational Testing Service.

Bates, D., & Maechler, M. (2017). *Matrix: Sparse and Dense Matrix Classes and Methods*. Recuperado de cran.rproject.org/web/packages/Matrix/Matrix.pdf

Bennet, D.C. (2001). *Evaluación de la calidad en la educación especial*. Londres: CNAA.

Betebenner, D. (2009). Growth, Standards and Accountability. *Research Archive Summary,* 1-21.

Betebenner, D., & Linn, R. (2010). Growth in Student Achievement: Issues of Measurement, Longitudinal Data Analysis, and Accountability, Educational Testing Service*. Measurement Challenges Within the Race to the Top Agenda*, 1-21.

Bethell, G. (2005). *Value-Added Indicators of School Performance: The English Experience Anglia Assessment*. Battisford: Suffolk, Inglaterra.

Bickel, R. (2007). *Multilevel analysis for applied research: It's just regression.* New York: Guilford.

Blanco, A., González, C., & Ordoñez, X. (2009). Patrones de correlación entre medidas de rendimiento escolar en evaluaciones longitudinales. *Revista de educación, 348,* 195-216.

Bloom, B. (1972*). Taxonomía de los objetivos de la educación; la clasificación de metas educativas*. Buenos Aires: El Ateneo.

Boscardin, C. K., Muthén, B., Francis, D. J., & Baker, E. L. (2008). Early identification of reading difficulties using heterogeneous developmental trajectories*. Journal of Educational Psychology, 100* (1), 192–208.

Boyle, M.H., & Willms, J.D. (2001). Multilevel modeling of hierarchical data in developmental studies. *Journal of Child Psychology and Psychiatry, 42*, 141-162.

Braun, H. (2005). Value-Added Modelling: What Does Due Diligence Require? En R. Lissitz (Ed), *ValueAdded Models in Education: Theory and Applications* (pp. 45-57). JAM Press, Maple Grove, MN.

Braun, H.., Wang, A., Jenkins, F., & Weinbaum, E. (2006). The Black-White achievement gap: Do state policies matter? *Education Policy Analysis Archives, 14* (8). Recuperado de epaa.asu.edu/epaa/v14n8/

Brian, R., Bill, V., Douglas, M., Kurt, H., Albrecht, G., & David, F. (2017). *MASS: Modern Applied Statistics with R*. Recuperado de cran.r-project.org/web /packages/MASS/MASS.pdf

Briggs, D. C. (2008). The Goals and Uses of Value-Added Models*. Washington: Commitee on Value-Added Methodology for Instructional Improvemente*. National Research Council and the National Academy of Education.

Briggs, D., & Betebenner, D. (2009). Is Growth Being Student Achievement Scale Dependent? *Annual meeting of the National Council for Measurement in Education*. San Diego CA: Academic Press.

Briggs, D., & Weeks, J. (2009). The Impact of Vertical Scaling Decisions on Growth Interpretations. *Educational Meaurement: Issues and Practice, 28* (4), 3-14.

Briggs, D., Weeks, J., & Wiley, E. (2008). Vertical Scaling in Value-Added Models for Student Learning*. At the National Conference on Value-Added Modeling*. Madison: WI.

Brown, S. (1994). School effectiveness research and the evaluation of schools. *Evaluation and Research in Education, 8*, 55-68.

Bryk, A. S. (1977). *An investigation of the effects of alternative statistical adjustment strategies in the analysis of quasi-experimental growth data* (doctoral thesis). Harvard University Graduate School of Education, Cambridge, MA.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 10,* 147-158.

Bryk, A.S., & Raudenbush, S.W. (1988). Toward a more appropriate conceptualization of research on school effects: A three-level hierarchical linear model. En R.D. Bock (Ed.), *Multilevel analysis of educational data* (pp. 159-204). San Diego: CA: Academic Press.

Bryk, A.S., & Raudenbush, S.W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park: SAGE.

Bryk, A. S., & Raudenbush, S. W. (2002). *Hierarchical Linear Models. Aplications and data analysis methods.* California: Sage Publications

Bryk, A.S., & Weisberg, H.I. (1976). Value-added analysis: a dyna- mic approach to the estimation of treatment effects. *Journal of Educational Statistics, 1* (2), 127-155.

Burstein, L. (1980). The analysis of multilevel data in educational research and evaluation. En D.C. Berliner (Ed.), *Review of research in education* (vol.8). Washington, DC: American Educational Research Association.

Calero, J., & Choi, A. (2012). La evaluación como instrumento de política educativa. *Presupuesto y gasto público, 67*, 29-42.

Campbell, D., & Kenny, D. (1999). *A primer on regression artifacts*. New York: Guilford Press.

Carabaña, J. (1987). Origen social, inteligencia y rendimiento académico al final de la EGB. En C. Lerena (ed.), *Educación y Sociología en España* (pp. 262-289). Madrid: Akal

Carlson, J. E. (2011). Statistical Models for Vertical Linking. En A. A. von Davier, *Statistical Models for Test Equating, Scaling, and Linking,* (pp. 59-70). New York: Springer.

Castellano, K.E., & Ho, A.D. (2013). Contrasting OLS and Quantile Regression Approaches to Student ''Growth'' Percentiles. *Journal of Educational and Behavioral Statistics, 38* (2), 190-215.

Castro. M. (1997). *Meta-análisis. Aportaciones metodológicas a la síntesis cuantitativa de la evidencia. Un estudio de simulación Montecarlo* (tesis doctoral). Universidad Complutense de Madrid.

Castro, M. (2014). 50 años del movimiento de "escuelas eficaces": lecciones aprendidas para el siglo XXI. *Participación Educativa, 3*(4), 9-16.

Castro, M. & Gaviria, J. L. (2009). Concepto y evolución de los modelos de valor añadido en educación. *Revista de Educación, 348*, 15-46.

Castro, M., Ruíz, C., & López, E. (2009). Forma básica del crecimiento en los modelos de valor añadido: vías para la supresión del efecto de regresión. *Revista de Educación, 348*, 111-136.

Chin, T., Kim, W., & Nering, M. (2006). Five Statistical Factors That Influence IRT Vertical Scaling. *Annual meeting of National Council on Measurement in Education (NCME)*. San Francisco.

Choi, K., Goldschmidt, P., & Yamashiro, K. (2006). *Exploring models of school performance form theory to practice.* University of California, CRESST, CSE. Los Angeles: University of California.

Choi, K., Seltzer, M., Herman, J., & Yamashiro, K. (2007). Children Left Behind in AYP and Non-AYP Schools: Using Student Progress and the Distribution of Student Gains to Validate AYP. *Educational Measurement: Issues and Practice, 26* (3), 21-32.

Chomsky, N. (2006). *Language and mind* (3rd ed.). Cambridge: Great Britain: Cambridge University Press.

Chung, H., & Beretvas, S.N. (2012). The impact of ignoring multiple membership data structures in multilevel models. *British Journal of Mathematical and Statistical Psychology, 65,*185–200.

Clarke, S., Neill, D., & Welbourn, R. (1960). The effects of corticotrophin and corticoids on secretion from denervated gastric pouches in dogs. *Gut, 1*, 36-43.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office.

Collins, L. M., & Horn, J. L. (1991). *Best methods for the analysis of change: Recent advances, unanswered questions, future directions.* Washington, DC: American Psychological Association.

Collins, L. M., & Sayer A. (2001). *New methods for the analysis of change*. Washington. D.C: American Psychological Association, APA.

Cook, T., & Campbell, D. (1979) *Quasi-experimentation: Design and analysis issues for field settings.* Boston: Houghton Mifflin.

Cook, N. R., & Ware, J. H. (1983). Design and analysis methods for longitudinal research. *Annual Review of Public Health, 4*, 1-23.

Council of Chief State School Officers, CCSSO (2008). *Implementer's guide to growth models:* Washington, DC: The Council of Chief State School Officers. Recuperado de www.ccsso.org/content/pdfs/IGG%20Final%20AP.pdf

Cowles, M., Carlin, B., & Connett, J. (1996). Bayesian Tobit modeling of longitudinal ordinal clinical trial compliance data with nonignorable missingness. *Journal of American Statistician Association, 91*, 86–98.

Crocker, L., & Algina, J. (1986). *Introduction to Classical* & Modern test theory. New York: Holt, Rinehart and Winston.

Cronbach, L. J. (1976). *Research on classrooms and schools: Formulation of questions, design, and analysis*. Stanford, CA: Stanford Evaluation Consortium.

Cronbach, L. J., & Webb. (1975). Between and within class effects in a reported aptitude by treatment interaction: Reanalysis of a study by G.L. Anderson. *Journal of Educational Psychology, 6,* 717-724.

Curran, P. J., Lee, T., Howard, A. H., Lane, S., & MacCallum, R. (2012). Disaggregating within-person and between-person effects in multilevel and structural equation growth models. *Longitudinal Methodology Social Behaviour*, 217-253.

D'Angiulli, A., Siegel, L. S., & Maggi, S. (2004). Literacy instruction, SES, and word-reading achievement in English language learners and children with English as a first language: A longitudinal study. *Learning Disabilities Research & Practice, 19,* 202–213.

Daniel, L. H. (2012). *Comparing cross-classified growth models with and without the cumulative effect of teachers to a hierarchical growth model on cross-classified data.* Pittsburgh: University of Pittsburgh.

De la Orden, A. (1985). Hacia una conceptualización del producto educativo. *Revista de Investigación Educativa, 3* (6), 271-283.

De la Orden, A. (2011). Reflexiones en torno a las competencias como objeto de evaluación en el ámbito educativo. *Revista Electrónica de Investigación Educativa*, *13* (2), 1-21.

De Leeuw, J. & Kreft, G. (1896). Random Coefficient Models for Multilevel Analysis*. Journal of Educational Statistics, 11*, 57-85.

Dedrick, R. F., Ferron, J. M., Hess, M. R., Hogarty, K. Y., Kromrey, J. D., & Lee, R. (2009). Multilevel modeling: a review of methodological issues and applications. *Review of Educational Research, 79*, 69–102.

Demie, F. (2003). Using Value-added data for school self-evaluation: a case study of practice in inner-city schools. *School Leadership & Management, 23* (4)*,* 445-467.

Denyer, M., Furnémount, J., Poulain, R., & Vanloubbeeck, G. (2007). *Las competencias en educación*. Un balance. México: Fondo de cultura económica.

Diaz-Emparanza. I. (1995). Selección del número de replicaciones en un estudio de simulación. *Estadística Española, 37* (140), 497-509.

Doran, H. C. (2003). Adding value to accountability. *Educational leadership*, *61* (3), 55–59.

Doran, H. C., & Izumi, L. T. (2004). *Putting Education to the Test: A Value-Added Model for California*. Recuperado de www.heartland.org/custom/semod_policybot/pdf/15626.pdf

Dorans, N. J. (2000)*. Distinctions among classes of linkages*. College Board Research Notes (RN-11). New York: The College Board.

Dorans, N. J. (2004). Equating, Concordance, and Expectation. *Applied Psychological Measurement, 28,* 227-246.

Dorans, N. J., Moses, T. P., & Eignor, D. (2011). Equating Test Scores: Toward Best Practices. En A. A. von Davier, Statistical Models for Test Equating, *Scaling, and Linking,* (pp. 21-42). New York: Springer.

Dorans, N. J., Pommerich, M., & Holland, P. W. (2007). *Linking and Aligning Scores and Scales*. New York: Springer.

Drury, D., & Doran, H. (2003). The Value of Value-Added Analysis. *NSBA Policy Research Brief, 3*, 1-4.

Eraut, M. (1994). *Developing professional knowledge and competence*. London: Routledge Falmer.

Expósito, E. (2015). *Efecto de la interración en el muestreo de sujetos e ítems sobre el error de enlace* (tesis doctoral). Universidad Complutense de Madrid.

Feng, L., Figlio, D., & Sass, T. (2010). School Accountability and Teacher Mobility. *CALDER*. Washington DC. Recuperado de www.urban.org/uploadedpdf/ 1001396-school-accountability.pdf

Ferrão, M. E. (2009). Sensibilidad de las especificaciones del modelo de valor añadido: midiendo el estatus socioeconómico. *Revista de Educación*, *348*, 137-152.

Fitz-Gibbon, C.T. (1997). *The value-added national project: Final report: feasibility studies for a national system of value added indicators.* London: School Curriculum and Assessment Authority.

Fitzmaurice, G., Laird, N., & Ware, J. (2004). *Applied longitudinal analysis*. New York: Wiley.

Franco, S. (2006). *The relationships among building level school/non school factors and valueadded scores in Ohio* (doctoral thesis). University of Cincinnati, Ohio.

Frederiksen, C. H., & Rotondo, J. (1979). Time series models and the study of longitudinal change. En J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in his study of behavior and development*. New York: Academic Press.

Fuchs, T., & Wossmann, L. (2007). What accounts for international differences in student performance? A re-examination using PISA data. *Empirical Economics, 32*, 433-464.

Furby, L. (1973). Interpreting regression toward the mean in developmental research. *Developmental Psychology, 8*, 172–179.

Galton, F. (1886). Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and Ireland, 15*, 246–263.

Galton, F. (1988). Regression towards mediocrity in hereditary stature. *Anthropological miscellanea, 246-263*.

Garson, D. (2013). *Longitudinal Analysis*. California: Statistical Associates.

Gaviria, J. L. (1989). *Proyecto Docente*. Madrid: Universidad Complutense (inédito).

Gaviria, J. L., Biencinto, C., & Navarro, E. (2009). Invarianza de la estructura de covarianzas de las medidas de rendimiento académico en estudios longitudinales en la transición de Educación Primaria a Secundaria. *Revista de Educación, 348*, 153-173.

Gaviria, J. L., & Castro, M. (2005). *Modelos jerárquicos lineales*. Madrid: La Muralla.

Gelman, A., Hesketh, S., Long, J., & Skrondal, A. (2008). *Analysis of Multivariate Social Science Data*. New York: CRC.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.

Gempp, R. (2010). Equiparación, alineamiento y predicción de puntuaciones en medición educativa. *Revista Iberoamericana de Evaluacion Educativa, 3* (2), 103-126.

Gianni R. (1964). *El libro de los errores.* París: Piergiorgio

Glass, G. V, Willson, V. L., & Gottman, J. M. (1972). *Design and analysis of lime-series experiments.* Boulder: University of Colorado, Laboratory of Educational Research.

Glenn, C., & De Groof, J. (2005). *Balancing Freedom, Autonomy and Accoutnability in Education.* Nijmegan, N.L: Wolf Legal Publishers.

Goldhaber, D. (2010). When the Stakes Are High, Can We Rely on Value-Added? Exploring the Use of ValueAdded Models to Inform Teacher Workforce Decisions. *Center for American* Progress Accessed. Recuperado de www.americanprogress.org/issues/2010/12/pdf/vam.pdf

Goldschmidt, P. (2002). *A comparison of student achievement in LAAMP and Non-LAAMP schools in Los Angeles County: Longitudinal analysis results 1997'98–1999-2000.* Los Angeles: University of California, Center for the Study of Evaluation (CSE).

Goldschmidt, P., Choi, K., & Martinez, F. (2004). *Using Hierarchical Growth Models to Monitor School Performance Over Time: Comparing NCE to Scale Score Results.* Los Angeles, CA: CSE.

Goldschmidt, P., Roschewski, P., Choi, K., Auty, W., Hebbler, S., Blank, R., & Williams, A. (2005). *Policymakers' Guide to Growth Models for School Accountabillity: How do Accountability Moels Differ.* Washington: CCSSO.

Goldstein, H. (1987). *Multilevel models in educational and social research*. New York: Oxford University Press.

Goldstein, H. (1988). Comparing schools. En H. Torrance (Ed.). *National assessment and testing: A research response*. London: BERA.

Goldstein, H. (1991). Better ways to compare schools? *Journal of Educational Statistics, 16* (2), 89-92.

Goldstein, H. (1997). Methods in School Effectiveness Research. *School Effectiveness and School Improvement, 8* (4), 369-395.

Goldstein, H. (2003). *Multilevel statistical models*. London: Arnold.

Goldstein, H., Burgess, S., & McConnell, B. (2007). Modelling the effect of pupil mobility on school differences in educational achievement. *Journal of the Royal Statistical Society, 170,* 941–954.

Goldstein, H. & deStavola, B. (2010). Statistical modelling of repeated measurement data *Longitudinal and Life Course studies, 1,* 170-185

Goldstein, H., & Spiegelhalter, D. J. (1996). League Tables and Their Limitations: Statistical Issues in Comparisons of Institutional Performance. *Journal of the Royal Statistical Society, 159* (3), 385-443.

Goldstein, H., & Woodhouse, G. (2001). Modelling Repeated Measurements. En A. H. Leyland., & H. Goldstein, *Multilevel Modelling of Health Statistics* (pp. 13-26). Chichester: John Willey and Sons.

Gong, B., Perie, M., & Dunn, J. (2006). *Using student longitudinal growth measures for school accountability under No Child Left Behind: An update to inform design decisions*. Recuperado de www.nciea.org

Gray, J., Jesson, D., Goldstein, H., Hedger, K., & Rabash, J. (1995). A multilevel analysis of school improvement: Changes in schools' performance over time. School. *Effectiveness and School Improvement, 6* (2), 97-114.

Grenn, J. L. (2010). *Estimating Teacher Effects Using Value-Added Models.* Lincoln: Dissertations and Theses in Statistics. University of Nebraska. Recuperado de digitalcommons.unl.edu/statisticsdiss/6

Guarino, C. M., Maxfield, M., Reckase, M. D., Thompson, P. N., & Wooldridge, J. M. (2015). An evaluation of empirical Bayes's estimation of value-added teacher performance measures. *Journal of Educational and Behavioral Statistics, 40* (2), 190-222.

Guttman, L. (1954). A New Approach to Factor Analysis: The Radex. En P. F. Lazarsfeld (Ed.), *Mathematical Thinking in the Social Sciences* (pp. 258-348). New York: Free Press.

Haebara, T. (1980). Equating logistic ability scales by a weighted least squared method. *Japanese Psychological Research, 22,* 144-149.

Haegeland, T., & Kirkeboen, L. (2008). *School performance and Value-Added indicators - What is the importance of controlling for socioeconomic background? A simple empirical illustration using Norwegian data.* Statistics:Norway.

Hajivassiliou, V. (1994). A simulation estimation analysis of the external debt crises of developing countries. *Journal of Applied Econometrics, 9*, 109–131.

Hambleton, R. K., & Swaminathan, H. (1987). *Item Response Theory: Principles and applications*. Boston, MA: Kluwer.

Hanushek, E. (1971). Teacher Characteristics and Gains in Student Achievement: Estimation usig micro data. *The American. Economic Review, 61* (2), 280-288.

Hanushek, E. (1972). *Education and Race*. Lexington, MA: D.C. Heath and Company.

Hanushek, E. (2003). The Failure of Input-Based Schooling Policies. *The Economic Journarl, 113,* 64-98.

Hanushek, E.A., & Raymond, M.E. (2004). The effect of school accountability systems on the level and distribution of student achievement. *Journal of the European Economic Association, 2* (2–3), 406–415.

Harris, C. W. (1967). *Problems in measuring change.* Madison: University of Wisconsin Press.

Harris, D. J. (2007). Practical Issues in Vertical Scaling. En N. J. Dorans, M. Pommerich, & P. W. *Holland, Linking and Aligning Scores and Scales* (pp. 233-251). New York: Springer.

Harvey, L. (2002). *La garantía de calidad en la educación superior*. Oslo: CNNA.

Harville, D. (1990). BLUP (Best Linear Unbiased Prediction) and Beyond. En D. Gianola & K. Hammond (eds), *Statistical Methods for Genetic Improvement of Livestock* (pp. 239-276). Springer, New York.

Harwell, M., Stone, C., Hsu, T., & Kirisci, L. (1996). Monte Carlo Studies in Item Response Theory. *Applied Psychological Measurement, 20* (1), 101-125.

Heck, R. H. (2000). Examining the impact of school quality on school outcomes and improvement: A value-added approach. *Educational Administrati Quarterly, 36* (4), 513-552.

Heck, R. H., & Thomas, S. (2009). *An Introduction to Multilevel Modeling Techniques*. New York: Routledge.

Henderson, R. (1950). Estimation of genetic parameters. *Ann. Math. Statist, 21*, 309-310.

Herrmann, M., Walsh, E., Isenberg, E., & Resch, A. (2013). Shrinkage of Value-Added Estimates and Characteristics of Students with Hard-to-Predict Achievement Levels. *Mathematica Policy Research, 17, 1-25.*

Hershberg, T., Simón, V. A., & Lea-Kruger, B. (2004). Measuring What Matters. *American School Board Journal. National School Boards Association*. Recuperado de www.asbj.com/2004/02/0204asbjhershberg.pdf

Hibpshman, T. (2004). A review of Value-Added Models. *Kentacky education Professional Standards Board.* Recuperado de www.kyepsb.net/documents/Stats/ Journals/Heterogeneity%20of%20regression.pdf

Hill, P. W. & Goldstein, H. (1998). Multilevel Modeling of Educational Data With Cross-Classification and Missing Identification for Units. *Journal of Educational and Behavioral Statistics, 23* (2), 117-128.

Holland, P. W. (2007). A framework and history for score linking. En N. J. Dorans, M. Pommerich, P. W. Holland, N. J. Dorans, M. Pommerich, & P. W. Holland (Eds.), *Linking and Aligning Scores and Scales* (pp. 5-30). New York, NY: Springer-Verlag.

Holland, P. W., & Dorans, N. J. (2006). Linking and Equating. En R. L. Brennan & R. L. Brennan (Ed.), *Educational Measurement* (4 ed., pp. 187-220). Westport, CT: Praeger Publishers.

Holland, P. W., Dorans, N. J., & Petersen, N. S. (2007). Equating Test Scores. In C. R. Rao & S. Sinharay, *Handbook of statistics 26. Psychometrics* (pp. 169-203). Amsterdam: North Holland.

Holland, P. W., & Hoskens, M. (2003). Clasical test theory as a first-order ítem response theory: application to true-scores prediction from a possibly nonparalel test. *Psychometrika, 68* (1), 123-149.

Horne, R. (2000). The performance of males and females in school and tertiary education. *Australian Quarterly, 72* (5/6), 21-26.

Hox J. (2002). *Multilevel Analysis: Techniques and Applications*. Mahwah, NJ: Erlbaum.

Hox, J. (2010). *Multilevel Analysis: Tecniques and Applications*. New York: Routledge.

Humphreys, L. G. (1960). Investigations of the simplex. *PsychoDielrika, 25*, 313-323.

Humphreys, L. G. (1968). The fleeting nature of the prediction of college academic success. *Journal of Educational Psychology, 59* (5), 375-380.

Instituto de Evaluación. (2009). *ICCS 2009. Estudio Internacional de Civismo y Ciudadania*. Madrid: Ministerio de Educación. Recuperado de www.madrid.org/cs/Satellite?blobcol=urldata&blobheader=application%2Fpdf &blobheadername1=Contentdisposition&blobheadername2=cadena&blobhead

ervalue1=filename%3Dinforme+espa%C3%B1ol+2009.pdf&blobheadervalue2=language%3Des%26site%3DPortalEducacion&blobkey=id&blobtable=MungoBlobs&blobwhere=1352835589301&ssbinary=true

Jakubowski, M. (2008). *Implementing Value-Added Models of School Assesment*. San Domenico di Fiesole: RSCAS.

Jamieson, J. (1995). Measurement of Change and the Law of Initial Values: A Computer Simulation Study. *Educational and Psychological Measurement, 55* (1), 38-46.

Jamieson, J. (1999). Dealing with baseline differences: two principles and two dilemmas. *International Journal of Psychophysiology, 31*, 155-161.

Jenewein, K., Knauth, P., & Zülch, G. (2002) *Kompetenzentwicklung I Unternehmensprozessen*. Aachen: Shaker.

Jiménez, J. (29 de marzo del 2017). Trump acaba con las reglas que obligaban a las escuelas a rendir cuentas al gobierno federal. *Univision*. Recuperado de www.univision.com/noticias/educacion/trump-acaba-con-las-reglas-que-obligaban-a-las-escuelas-de-entrenamiento-de-profesores-a-rendir-cuentas-al-gobierno-federal

Jong, F. (2003). Leren in en voor een beroep. En M. Mulder, R. Wesselink, H. Biemans, L. Nieuwenhuis & R. Poell (Eds), *Compentiegericht beroepsonderwijs. Gediplomeerd, maar ook bekwaam?* (pp. 234-26). Wolters Noordhoff, Houten.

Joreskog, K. G. (1970). Estimation and testing of simplex models. *British Journal of Mathematical and Statistical Psychology, 23*, 121-145.

Joreskog, K. G., & Sorbom, D. (1977). Statistical models and methods for analysis of longitudinal data. En D. J. Aigner & A.S. Goldberger (Eds.), *Laten variables in socio-economic models*. Cambridge. MA: Abt Books.

Jungnam, K. (2007). *A Comparison of Calibration Methods and Proficiency Stimators for Creating IRT Vertical Scales* (doctoral thesis). Recuperado de ir.uiowa.edu/etd/163

Kaczynska, M. (1965). *El rendimiento escolar y la inteligencia*. Madrid: Espasa Calpe.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80,* 237–251.

Kane, T. J., & Staiger O. (2002). The Promise and Pitfalls of Using Imprecise School Accountability Measures. *Journal of Economic Perspectives, 16* (4), 91–114.

Kelley, T. (1947). *Fundamentals of statistics.* Cambridge, MA: Harvard University.

Kenny, D. A. (1979). *Correlation and causality*. New York: Wiley-Interscience.

Khoo, S., West, S. G., Wu, W., & Kwok, O. (2006). Longitudinal methods. En M. Eid & E. Diener (Eds.), *Handbook of psychological measurement: A multimethod perspective* (pp. 301–317). Washington, DC: APA.

Koedel, C., & Betts, J. R. (2009). *Does Student Sorting Invalidate Value-Added Models of Teacher Effectiveness? An Extended Analysis of the Rothstein Critique. Working* Papers 0902. Department of Economics, University of Missouri. Recuperado de economics.missouri.edu/working-papers/2009/wp0902_koedel.pdf

Kolen, M. J. (2004). Linking Assessments: concept and history. *Applied Psychological Measurement, 28* (4), 219-226.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and Practices.* New York: Springer.

Kolen, M. J., & Brennan, R. L. (2014). *Test Equating, Scaling, and Linking. Methods and Practices. Third Edition.* New York: Springer.

Krause, A., & Pinheiro, J. (2007). Modeling and simulation to adjust p values in presence of a regression to the mean effect. *The American Statistician, 61,* 302–307.

Kreft, I. G., & De Leeuw, J. (1998). *Introducing multilevel modeling*. Newbury Park, CA: Sage.

Kupermintz, H. (2003). Teacher effects and teacher effectiveness: A validity investigation of the Tennessee value added assessment system. *Educational evaluation and policy analysis, 25* (3), 287–298.

Lacey, J., & Lacey, B. (1962). The law of initial value in the longitudinal study of autonomic constitution: reproductibility of autonomic responses and response patterns over a four-year interval. En W. Wolf (ed.), *Rhythmic functions in the living system* (pp. 1257-1290). Annals of the New York Academy of Science.

Ladd, H., & Walsh, R. (2002). Implementing value-added measures of school effectiveness: getting the incentives right. *Economics of Education Review , 21*, 1-17.

Laird, N., & Ware, J. (1982). Random-effects Models for Longitudinal Data. *Biometrics 38*, 963-974.

Lee, V. (2000). Using hierarchical linear modeling to study social contexts: The case of school effects. *Educational Psychologist, 35,* 125–141.

Lee, W. C., & Ban, J.-C. (2010). A Comparison of IRT *Linking Procedures. Applied Measurement in Education, 23* (1)*,* 23-48.

Lee, M., & Smith, G. (2002). Regression to the mean and football wagers. *Journal of Behavioral Decision Making, 15,* 329–342.

Ley Orgánica 8/2013, de 9 de diciembre, *para la Mejora de la Calidad Educativa*. BOE

Linn, R. L. (1980). Discussion: Regression toward the mean and the interval between test administrations. *New Directions for Testing and Measurement, 8*, 83–89.

Linn, R. L. (1993). Linking results of distinct assessments. *Applied Measurement Education, 6* (1), 83-102.

Linn, R. L. (2005). Conflicting demands of No Child Left Behind and state systems: Mixed messages about school performance. *Education Policy Analysis Archives, 13*(33)*.* Recuperado de epaa.asu.edu/epaa/v13n33/

Linn, R. L., & Haug, C. (2002). Stability of school-building accountability scores and gains. *Educational Evaluation and Policy Analysis, 24* (1)*, 29-36.*

Lissitz, R. W., Doran, H., Schafer, W. D., & Willhoft, J. (2006). Growth Modeling, Value Added Modeling and Linking: An introduction. En R. W. Lissitz (Ed), *Longitudinal and Value Added Models of Student Performance* (pp. 1-46)*.* Minnesota: JAM Press.

Little, R., Milliken, G., Stroup, W., Wolfinger, R., & Schabenberger, O. (2007). *SAS for Mixed Models*. USA: SAS.

Little, R., & Rubin, D. (1987). *Statistical analysis with missing data*. New York: Wiley.

Little, T., Schnabel, K., & Baumert, J. (Eds.). (2000). *Modeling longitudinal and multilevel data*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Lizzio, A., & Wilson, K. (2004). Action learning in higher education: an investigation of its potential to develop professional capability. *Studies in Higher Education, 29,* 469–488.

Lockwood, J. R., Louis, T. A. & McCaffrey, D. F. (2003). Uncertainty in Rank Estimation: Implications for Value-Added Modeling Accountability. *Journal of Educational and Behavioral Statistics, 27* (3), 255-270.

Lockwood, J. R., & McCaffrey, D.F. (2009). Exploring Student-Teacher Interactions in Longitudinal Achievement Data. *Education Finance and Policy, 4* (4), 439-467.

Lockwood, J. R., McCaffrey, D. F., Hamilton, L. S., Stecher, B., Le, V-N., & Martinez, J. F. (2007). The Sensitivity of Value-Added Teacher Effect Estimates to Different Mathematics Achievement *Measures. Journal of Educational Measurement, 44 (1)*, 47–67.

Lonford, N. (1989). To center or not to center. Multilevel *Modelling Newsletter, 1* (3), 7-11.

Lonford, N. (1993). *Random coefficient models*. New York: Oxford University Press.

Lord, F. M. (1956). The measurement of growth. *Educational and Psychological Measurement*, *16,* 421–437.

Lord, F. M. (1958). Further problems in the measurement of growth. *Educational and Psychological Measurement, 18,* 437-454.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. New Jersey: Lawrence Erlbaum associates, publishers.

Lord, F., & Novick, M. (1968). *Statistical theory of mental test scores.* Addison-Wesley: Reading, MA.

Loyd, B., & Hoover, H. D. (1980). Vertical equating using the Rasch model. *Journal of Educational Measurement, 17,* 179-193.

Lunneborg, C. E., & Lunneborg, P. W. (1970). Relations between aptitude changes and academic success during college. *Journal of Educational Psychology, 61* (3), 169-173.

Luo, W., & Kwok, O. (2009). The impacts of misspecifying cross-classified random effects models. *Multivariate Behavioral Research, 44*, 182–212.

Luo, W., & Kwok, O. (2012). The consequences of ignoring individual's mobility in multilevel growth models: A Monte Carlo study. *Educational and Behavioral Statistics, 37,* 31–56.

Macbeath, J., & Mortimore, P. (2001). *Improving school effectiveness*. Buckingham: Open University Press.

Marchesi, A., Martínez, R. & Martín, E. (2004). Estudio longitudinal sobre la influencia del nivel sociocultural en el aprendizaje de los alumnos en la Educación Secundaria Obligatoria. *Infancia y Aprendizaje, 27* (3), 307-323.

Marco, G. L. (1977). Item characteristic curve solutions to three intractable testing problems. *Journal of Educational Measurement, 14*, 139-160.

Marsh, H. & Hau. K. (2002). Multilevel Modeling of Longitudinal Growth and Change: Substantive Effects or Regression Toward the Mean Artifacts? *Multivariate Behavioral Research*, *37* (2), 245-282.

Martineau, J. (2009). Measuring Student Achievement Growth at the Hisgh School Level. En L. Pinkus, *Meaningful Measurement: The Role of Assessments in Improving High School Education in the Twenty-Frist Century* (pp. 119-142). Washington: Alliance for Excellent Education.

Martínez-Arias, R. (2009). Usos, aplicaciones y problemas de los modelos de valor añadido en educación. *Revista de Educación, 348*, 217-250.

Martínez-Arias, R., Gaviria, J. L., & Castro, M. (2009). Concepto y evolución de los modelos de valor añadido en educación. *Revista de Educación, 348*, 15-47.

Martínez-Otero, V. (1996). *Factores determinantes del rendimiento académico en la enseñanza Media.* Madrid: Publicaciones del Colegio de Psicólogos de Madrid.

Mass, C., & Hox, J. (2005). Sufficient Sample Sizes for Multilevel Modeling. Methodology, *1* (3), 86–92.

Matthew, M. (2010). Modeling Outcomes with Floor or Ceiling Effects: An Introduction to the Tobit Model. *Gifted Child Quarterly, 54* (4), *314–320*.

McCaffrey, D. F., Koretz, D., Louis, T. A., & Hamilton, L. (2004). Models for Value-Added Modeling of Teacher Effects. *Journal of Educational and Behavioral* Statistics, *29* (1), 67-101.

McCaffrey, D. F, Lockwood, J. R, Doretz, D., & Hamilton, L. (2003a). *Evaluating Value-Added Models for Teacher Accountability.* Santa Monica: RAND Corporation.

McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., Louis, T. A., & Hamilton, L. (2003b). Models for value-added modelling of teacher effects. *Journal of Education and Behavioral Statistics*, *29* (1), 67-101.

Mccall, M. S., Kingsbury, G. G., & Olson, A. (2004). *Individual growth and school success*. Lake Oswego, OR: Northwest Evaluation Association. Recuperado de http://www.nwea.org

McClelland, D. C. (1973). Testing for Competence Rather Than for "Intelligence". *American Psychologist, 28,* 1-14.

McClelland, D.C. (1998). *Identifying competencies with behavioural-event interviews. Psychological Science, 9* (5), 331-339.

McGeevor, P., Giles, C., Little, B., Head, P., & Brennan, J. (1990). *The measurement of Value Added in higher education*. A joint PCFC/CNAA project report.

McNamara, J., Scissons, M., & Gutknecth, N. (2011). A longitudinal study of kindergarten children at risk for reading disabilities: The poor really are getting poorer. *Journal of Learning Disabilities, 44* (4), 421-430.

McPherson, A. (1992). *Measuring value added in schools (NCE briefing 1).* London: National Commission on Education.

MECD, Ministerio de Educación, Cultura y Deporte. (2015). Relaciones entre las competencias, los contenidos y los criterios de evaluación de la educación primaria, la educación secundaria obligatoria y el bachillerato. Madrid: Ministerio de Educación, Cultura y Deporte.

Merriënboer, J., Van der Klink, M., & Hendriks, M. (2002). *Competenties: van complicaties tot compromis. Over schuifjes en begrenzers.* Den Haag: Onderwijsraad.

Merton, R. (1968). The Matthew Effect in Science. En R. Merton, *The Sociology of Science*. (pp. 56-63). Chicago: Norman W. Storer.

Merton, R. (1988). The Matthew Effect in Science, II Cumulative Advantage and the Symbolism of Intellectual Property. *The History of Science Society, 79* (4), 606-623.

Messick, S. (1984). The psychology of educational measurement. *Educational Measurement, 21*, 215–237.

Meyers, J. L., & Beretvas, S. N. (2006). The Impact of Inappropriate Modeling of Cross-Classified Data Structures. *Multivariate Behavioral Research, 41* (4), 473–497.

Miller, S. A. (1998). *Developmental research methods* (2nd ed.). Upper Saddle River, NJ: Prentice-Hall.

Mislevy, R. J. (1992). *Linking educational assessments: Concepts, issues, methods, and prospects*. Princeton, NJ: ETS Policy Information Center.

Morgan, L., Farkas. G., & Hibel.J. (2015). Matthew Effects for Whom? *Learn Disabil Q, 31* (4), 187–198.

Morgan, L., Farkas, G., & Wu, Q. (2011). Kindergarten children's growth trajectories in reading and mathematics: Who falls increasingly behind? *Journal of Learning Disabilities, 44* (4), 472-488.

Mulder, M. (2001). Competence development - some background thoughts, *Journal of Agricultural Education and Extension, 7* (4), 147-159.

Muñiz, J. (2003). *Teoría clásica de los tests*. Madrid: Ediciones Pirámide.

Muñiz, G., Ribitaille, A., Kelly, A., Johansson, B., Hofer, S., & Piccinin. A. (2017). Latent growth models matched to research questions to answer questions about dynamics of change in multiple processes*. Journal of Clinical Epidemiology, 82*, 159-166.

Muñoz, A. (1977). *Valoración del rendimiento de centros docentes de EGB. VI Plan de Investigación Educativa. CIDE (MEC).* Extremadura: ICE.

Murillo, F. J. (2003). Una panorámica de la investigación iberoamericana sobre eficiencia escolar. *Revista electrónica iberoamericana sobre calidad, eficiencia y cambio en educación, 1*, 1-14.

Murillo, F. J. (2008). Los modelos multinivel como herramienta para la investigación educative. *Magis, Revista Internacional de Investigación en Educación, 1*, 45-62.

Murnane, R. (1975). *The impact of school resources on the learning of inner city children*. Cambridge, MA: Ballinger.

Murphy. (2012). Where is the Value in Value-Added Modeling? *Pearson Education*, 1-14.

National Assessment of Education Progress. (1977). NAEP data explorer. NCES. Recuperado de nces.ed.gov/1nationreportcard/nde/viewresults

Navarro, E. (2013). *El Valor Añadido en Educación: Cuestiones Teóricas y Metodológicas* (tesis doctoral). Universidad Complutense de Madrid.

Nelson, C. R. (1973). *Applied time series analysis*. San Francisco: Holden-Day.

Nesselroade, J., Stigler, S., & Baltes, P. (1980). Regression toward the mean and the study of change. *Psychological Bulletin, 88*, 622-637.

Nijhof, W.J. (2003). Naar competentiegericht beroepsonderwijs? En M. Mulder, R. Wesselink, H. Biemans, L. Nieuwenhuis & R. Poell (Eds.), *Compentiegericht beroepsonderwijs. Gediplomeerd, maar ook bekwaam?* (pp. 120-135). Houten: Wolters Noordhoff.

O'Connell, A. & McCoach, B. (Eds.). (2008). *Multilevel modeling of educational data.* Charlotte, NC: Information Age.

OECD. (2008). *Measuring Improvements in Learning Outcomes.* París: OECD.

OECD. (2011). *Education at a Glance 2011: OECD Indicators*. OECD Publishing

OECD. (2016). *Average class size*. París: OECD. Recuperado de www.oecd.org/about/

Ogasawara, H. (2002). Stable response functions with unstable item parameter estimates. *Applied Psychological Measurement, 26* (3), 239-254.

Pacheco, B. (1970). Aprovechamiento escolar y aptitudes mentales. *Revista Educadores, 59*, 539-554.

Pardo, M., & Ruiz, M. (2013). *Análisis de Datos en Ciencias Sociales y de la Salud III*. Madrid: Síntesis.

Patz, R. J. (2007). *Vertical Scaling in Standards-Based Educational Assessment and Accountability Systems*. Washington, DC: CCSSO.

Phillips, G., & Adcock, E. (1996). *Practical applications of hierarchical linear models to district evaluations.* Annual Meeting of the American Educational Research Association. New York.

Ponisciak, S. M., & Bryk, A. S. (2005). Value Added Analysis of the Chicago Public Schools: An application of Hierarchical Models. En L. R, *Value Added Models in Education: Theory and Applications* (pp. 40-79). Maple Grove, MN: JAM Press.

Protopapas, A., Sideridis, G., Mouzaki, A., & Simos, P. G. (2011). Matthew Effects in reading comprehension: Myth or reality? *Journal of Learning Disabilities, 44* (5), 402-420.

Psychometric Society. (1979). Publication policy regarding Monte Carlo studies. *Psychometrika, 44*, 133-134.

R Core Team. (2017). R: A language and enviorement for statistical computing. Foundation for statistical Computing, Vienna, Austria. Recuperado de www.r-project.org/

Ramirez, D., Yuen, R., Ramey, R., & Pasta, D. (1991). *The immersion study*. Washington: Office of Educational Research and Improvement.

Rao, C. R. (1965). The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. *Biometrika 52*, 447-458.

Raudenbush, S. W. (1993). A crossed random effects model for unbalanced data with applications in cross-sectional and longitudinal research. *Journal of Educational Statistics, 18,* 321–349.

Raudenbush, S. W. (2004). What are Value-Added models estimating and what does this imply for statistical practice? *Journal of Educational and Behavioral Statistics, 29* (1), 121-129.

Raudenbush, S. W., & Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociology of Education, 59*, 1-17.

Raudenbush S. W., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods*. United States: CA: Sage Publications.

Raudenbush, S. W. & Willms, J.D. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics, 20*, 307-335.

Ray, A. (2006). *School Value Added Measures in England: A Paper for the OECD Project on the Development of Value-Added Models in Education System.* London: Department for Education and Skills. Recuperado de www.dcsf.gov.uk/research /data/uploadfiles/RW85.pdf

Ray, A., Evans, H., & McCormack, T. (2009). El uso de los modelos nacionales de valor añadido para la mejora de las escuelas británicas. *Revista de Educación, 348*, 47-66.

Ray, A., Helen, E., & Tanya, M. (2009). School Value Added Measures in England. *Revista de Educación, 348*, 47-66.

Raykov, T. (1995). On statisitical approaches to the study of the «Law of Initial Values». *Quality and Quantity, 29*, 251-271.

Reardon, S. F., & Raudenbush, S. W. (2008). Assumptions of Value-Added Models for Estimating School Effects. *National Conference on Value-Added Modeling.* Madison: University of Wisconsin.

Reckase, M. D. (2008). *Measurement Issues Associated with Value-added Methods.* National Research Council Committee on Value-added Methodology. Washington, DC.

Reckase, M. D. (2010). *Study of best practices for vertical scaling and standard seting with recommendations for FCAT 2.0*. Recuperado de www.fldoe.org/ asp/k12memo/pdf/StudyBestPracticesVerticalScalingStandardSetting.pdf

Reisner, E., Alkin, C., Boruch, R., Linn, R., & Millman, J. (1982). *Assessment of the title I evaluation and reporting system*. Washington, D.C: U.S. Department of Education.

Reynolds, D., & Creemers, B. (1990). School Effectiveness and School Improvement: a Mission Statement. *School Effectiveness and School Improvement, 1*,1-3.

Richter, T. (2006). What's wrong with ANOVA and multiple regression? Analyzing sentence completion times with hierarchical linear models. *Discourse Processes, 41,* 221–250.

Roberts, A. (1980). Regression toward the mean and the regression-effect bias. En G. Echternacht (ed), *New directions for testing and measurement* (Vol. 8, pp. 59–82).

Robinson, W. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review, 15*, 351-357.

Robinson, G. (1991). That BLUP is a Good Thing: The Estimation of Random Effects**.** Statistical *Science 6* (1), 15-32.

Rocconi, L., & Ethington, C. (2009). Assessing Longitudinal Change: Adjustment for Regression to the Mean Effects. *Res High Education, 50,* 368–376.

Rogosa, D. R. (1979). Causal models in longitudinal research: Rationale, formulation and interpretation. En J. R. Nesselroade y P. B. Baltes (Eds.), *Longitudinal research in the study of behaviour and development* (pp. 263-302). New York: Academic Press.

Rogosa, D. R. (1988). Myths about longitudinal research. En K. W. Schaie, R. T. Campbell, W. Meredith & S. C. Rawlings (Eds.), *Methodological issues in aging research* (pp. 171-210). New York: Springer.

Rogosa, D. R. (1995). Myths and methods: Myths about longitudinal research, plus supplemental questions. En J. M. Gottman, *The analysis of change* (pp. 3-65). Lawrence Erlbaum Associates, New Jersey.

Rogosa, D. R., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin, 90,* 726-748.

Rogosa, D. R., & Willett, J. B. (1983). Demostrating the reliability of the difference score in the measurement of change. *Journal of Educational Measurement, 20* (4), 335-343.

Rogosa, D. R., & Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika, 50*, 203-228.

Ronald, H. H., Scott, L.T., & Lynn, N.T. (2010). *Multilevel and longitudinal modeling with SPSS*. New York: Routledge.

Rosário, M., Silva, F., Gazaffi, R., Moura, A., Ledur, M., Garcia, A., & Coutinho, L. (2011). *Modelos mistos na modelagem fenotípica da população F2 desenvolvida para mapeamento de QTLs na galinha. Avicultura*. Recuperado de pt.engormix.com/MA-avicultura/genetica/artigos/modelos-mistos-modelagem-fenotipica-t472/103-p0.htm

Rothstein, J. (2009). Student sorting and bias in value added estimation: Selection on observables and unobservables. *National Bureau of Economic Research*. Recuperado de www.princeton.edu/~ceps/workingpapers/ 170rothstein.pdf

Rovine, M., & VonEye, A. (1991). *Applied computational statistics in longitudinal research.* San Diego, CA: Academic Press.

Rowan, B., Correnti R., & Miller R. J. (2002). What large scale, survey research tells us about teacher effects on student achievement: Insights from the prospects study of elementary schools. *Teachers College Record, 104,* 1525–1567.

Rubin, D. (1976). Inference and missing data. *Biometrika, 63*, 581–592.

Rubin, D., Stuart, E., & Zanutto, E. (2004). A potential outcome view of value-added assessment in education. *Journal of Educational and Behavioral Statistics, 29* (1), 103-116.

Ruiz de Miguel, C., & Castro, M. (2006). Estudio multinivel basado en PISA 2003: factores de eficacia escolar en el área de matemáticas. *Education Policy Analysis Archives, 14* (29), 1-26.

Salvador, F., Rodríguez, L., & Bolivar, A. (2004). Diccionario *Enciclopédico de Didáctica* (Vol. II). Málaga: Aljibe.

Sanders, W. L. (2000). Value-added assessment from student achievement data: Opportunities and hurdles. *Journal of Personnel Evaluation in Education, 14* (4), *329–339.*

Sanders, W. L. (2006). Comparisons among various educational assessment value added models*. The Power of Two National Value-Added Conference*. Columbus, Ohio.

Sanders, W. L., & Horn, S. P. (1994). The Tenessee Value Added Assessment System (TVAAS): Mixed-model methodology in educational assessment*. Journal of Personnel Evaluation in Education, 8,* 299-311.

Sanders, W. L., & Horn, S. P. (1998). Research Findings from the Tennessee Value-Added Assessment System (TVAAS) Database: Implications for Educational Evaluation and Research. *Journal of Personnel Evaluation in Education, 12* (3), 247-256.

Sanders, W. L., Saxton, A. M., & Horn, S. P. (1997). The Tennessee Value-Added Accountability System: A Quantitative, Outcomes-Based Approach to Educational Assessment. En J. Millman (Ed), *Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure?* (pp. 137-162). Thousands Oaks: Corwin Press.

Sanders, W. L., & Wright, S. P. (2008). *A Response to Amrein-Beardsley (2008) "Methodological Concerns about the Education Value- Added Assessment System.* Recuperado de www.sas.com/govedu/edu/services/Sanders_Wright_response_to_Amrein-Beardsley_4_14_2008.pdf

Saunders, L. (1999). A brief history of educational «Value Added»: How did we get to where we are? *School Effectiveness and School Improvement, 10,* 233-256.

Scheerens, J. (2005). The quality imperative. Commissioned for the EFA Global Monitoring report 2005. *Review of school and instructional effectiveness research*.

Searle, S., Casella, G., & McCulloch, C. (1992). *Variance components.* New York: John.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Shin, T. (2007). Comparison of thee growth modeling techniques in the multilevel analysis of longitudinal academic achievement scores: Latent growth modeling, hierarchical linear modeling, and longitudinal profile analysis via multidimensional scaling. *Asia Pacific Education Review, 8* (2), 262–275.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and even occurrence.* New York: Oxford University.

Skaggs, G., & Lissitz, R. (1982). Test Equating: Relevant Issues and a Review of recent Research. *Anual Meeting of the American Educational Research Association*. Los Ángeles.

Snijders, T., & Bosker, R. (1999). *Multilevel Analysis*. London: Sage.

Snijders, T., & Bosker, R. (2012). *Multilevel Analysis: An introduction to basic and advanced multilevel modeling*, 2nd Edition. Thousand Oaks, CA: Sage.

Spours, K., & Hogdson, A. (1996). *Value Added and raising attainment: A formative approach.* A resource pack for practitioners. Poole: BP Education Service.

Stanovich, K. (1986). Matthew effects in reading: some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21,* 360-407.

Stanovich, K. (1993). *Does* reading make you smarter? Literacy and the development of verbal intelligence. *Advances in Child Development and Behaviour, 24, 133-180.*

Stevens, J. J. (2000). *Educational accountability systems: Issues and recommendations for New Mexico* (technical report). New Mexico State Department of Education.

Stevens, J. J. (2005). The Study of School Effectiveness as a Problem in Research Design. En R. W. Lissitz (Ed.), *Value Added Models in Education, Theory and Applications* (pp. 167-208). Maaple Grove, MN: JAM Press.

Stevens, J. J., & Zvoch, K. (2006). Issues in the implementation of longitudinal growth models for studen achievement. En R. W. Lissitz (Ed), *Longitudinal and Value Added Models of Student Performance* (pp. 170-209). Maple Grove: JAM Press.

Stocking, M. L. & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement* (7), 201-210.

Stone, J., & Norman, J. (1993). Numerical simulations of protostellar jets with nonequilibrium cooling. I - Method and two-dimensional results. *Astrophysical Journal, 413* (1), 198-220.

Tabachnick, B., & Fidell, L. (2007). *Using multivariatestatistics* (5th ed.). Needham Heights, MA: Allyn & Bacon.

Tallmadge, G. (1982). An empirical assessment of norm-referenced evaluation methodology. *Journal of Educational Measurement, 19*, 97–112.

Tate, R. (2004). A Cautionary Note on Shrinkage Estimates of School and Teacher Effects. *Florida Journal of Educational Research, 42*, 1- 21.

Taylor, J., & Nguyen, A.N. (2006). An Analysis of the Value Added by Secondary Schools in England: Is the Value Added Indicator of Any Value? *Oxford Bulletin of Economics and Statistics, 68*, 203-224.

Teddlie, C., & Reynolds, D. (2000). *The International Handbook of School Effectiveness Research*. New York: Falmer Press.

Tekwe, C., Carter, R., Ma, C., Lucas, M., Roth, J., Ariet, M., Fisher, T. & Resnick, M. (2004). An Empirical Comparison of Statistical Models for Value-Added Assessment of School Performance. *Jounal of Educational and Behavioral Statistics, 29* (1), 11-36.

Thiemea, C., Prior, D., Tortosa-Ausinac, E., & Gempp, R. (2016). Value added, educational accountability approaches and their effects on schools' rankings: Evidence from Chile. *European Journal of Operational Research, 253*, 456-471.

Thorndike, R. (1939). On the fallacy of imputing the correlations found for groups to the individuals or smaller groups composing them. *Am J Psychol, 52,*122-4

Thorndike, R. (1963). *The concepts of over and under achievement.* New York: Teachers College, Columbia University.

Thum, Y. M. (2003). *No Child Left Behind: Methodological challenges & recommendations for measuring adequate yearly progress (CSE Technical Report 590)*. Los Angeles: Center for Research on Evaluation, Standards, and Student Testing, UCLA.

Thum, Y. M. (2006). Designing Gross Productiviy Indicators: A Proposal for Connecting Accountability Goals, Data, and Analysis. En R. W. Lissitz, *Longitudinal and Value Added Models of Student Performance* (pp. 436-479). Maple Grove: JAM Press.

Thum, Y. M. (2009). No Child Left Behind: Retos metodológicos y recomendaciones para la medida del progreso anual adecuado. *Revista de Educación, 348*, 67-90.

Tobin, J. (1958). Estimation of relationship for limited dependent variables. *Econometrica, 26*, 24-36.

Tong, Y., & Kolen, M. J. (2007). Coparisons of Methodologiaes and Results in Vertical Scaling for Educational Achievement Test. *Applied Measurement in Education, 20* (2), 227-253.

UNESCO. (2016). *Education for people and planet: Creating sustainable futures for all* (Second edition). París: UNESCO.

Vallejo, S. G. (1989). Regresión de series de tiempo con mediciones igualmente espaciadas. *Anuario de psicología, 43*, 129-155.

Vallejo, S. G. (1990). Diseños de medidas repetidas con errores autocorrelacionados. *Psicothema, 2,* 189-209.

Von Davier, A. A. (2011). *Statistical Models for Test Equating, Scaling and Linking*. New York: Springer.

VonEye, A. (1990). *Statistical methods in longitudinal research: Volume I-Principles and structuring change.* Boston: Academic Press.

Wainer, H. (2004). Introduction to the value-added assessment special issues. *Journal of Educational and Behavioral Statistics, 29* (1)*,* 1-4.

Walberg, H., & Tsai, S. (1983). Matthew effects in education. *American Edzlcational Research Journal, 20*, 359-373.

Wallis, W., & Steptoe, S. (2007). How to fix No Child Left Behind? *Newsweek, 169* (23)*,* 34–41.

Wang, L., Zhang, Z., McArdle, J., & Salthouse, T. (2009). Investigating Ceiling Effects in Longitudinal Data Analysis. *Multivariate Behavioral Research, 43* (3), 476–496.

Webster, W. J. (2005). The Dallas School-Level Accountability Model: The Marriage os Status and Value-Added Approches. En R. W. Lissitz (Ed.), *Value-Added Models in Education. Theory and Applications* (pp. 233-271). Maple Grove: JAM Press.

Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y., & Zemla, J. (2017). Corrplot. a correlation matrix or general matrix. Recuperado de https://cran.rproject.org/web/ packages/corrplot/corrplot.pdf

Werts. C. E., & Hilton. T. L. (1977). Intellectual status and intellectual growth again. *American Educational Research Journal, 14,* 137-146.

Wilder, J. (1957). The lay of initial value in neurology and psychiatry. *Journal of Nervous and Mental Disease, 125*, 73-86.

Wiley, E. W. (2006). *A Practitioner's Guide to Value Added Assessment.* Tempe: Educational Policy Studies Laboratory, Arizona State University.

Willett, J. B. (1988). Questions and answers in the measurement of change. En E. Z. Rothkopf (Ed.), *Review of research in education* (Vol. 15, pp. 345-422)*.* Washington, DC: American Educational Research Association.

Willett, J. B. (1989). Questions and answers in the measurement of change. En E. Z. Rothkopf (Ed.), *Review of research in education* (Vol. 15, pp. 345-422)*.* Washington, DC: American Educational Research Association.

Willett, J. B. (1994). Measuring change more effectively by modeling individual change over time. En T. Husen & T. N. Postlethwaite (Eds.), The international encyclopedia of education (2nd ed.). Elmsford, NY: Pergamon Press.

Willett, J. B. (1997). Measuring change: what individual growth modelling buys you. En E. Amsel & K. A. Renninge (Eds.), *Change and Development: Issues of Theory, Method, and Application* (pp. 45-87). Mahwah, NJ: Lawrence Erlbaum Associates.

Willet, J. B.; Singer, J. D. y Martín, N. C. (1998). The design and analysis of longitudinal studies of development and psychopathology in context: Statistical models and methodological recommendations. *Developmental and Psychopathology, 10,* 395-426.

Willms, J. D. & Raudenbush, S. W. (1989). A longitudinal hierarchical linear model for estimating school effects and their stability. *Journal of Educational Measurement, 2,* 209-232.

Willms, J. D. (1992). *Monitoring school performance: A guide for educators*. Washington, DC: Falmer Press.

Winterton, J., Delamare Le-Deist, F., & Stringfellow, E. (2005). *Typology of knowledge, skills and competences: clarification of the concept and prototype*. Thessaloniki: Cedefop.

Wonnacott, T., & Wonnacott, R. (1999). *Introducción a la estadística*. México: Limusa.

Wright, S. P., Horn, S. P., & Sanders, W.L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11,* 57-67.

Wright, S. P., Sanders, W. L. & Rivers, J. C. (2006). Measurement of academic growth of individual students toward variable and meaningful academic standards. En R. Lissitz (Ed.), *Longitudinal and value added models of student performance.* Maple Grove, MN: JAM Press.

Wright, S. P., White. J. T., Sanders, W. L., & Rivers. J. C. (2010). SAS EVAAS Statistical Models. Cary, NC: SAS Institute. Recuperado de www.sas.com/ resources/asset/SAS-EVAAS-Statistical-Models.pdf

Yang, M., Goldstein, H., Rath, T., & Hill, N. (1999). The use of assessment data for school improvement purposes. *Oxford Review of Education, 25*, 469-483.

Yeagley, R. (2007). Separating Growth from Value Added: Two Academic Models Offer Different Tools for Different Purposes. *The School Administrator, 64* (1). Recuperado de www.aasa.org/ publications /saarticledetail.cfm? Item Number=7941

Younk, D. J. (1999). *Studen Progress in Australian Schools: A Multilevel Multivariate Model. Montreal*: Paper presented to the American Educational Research Association Annual Meeting.

Zapata, L., De Los Reyes, C., Lewis S., & Barceló, E. (2009). Memoria de trabajo y rendimiento académico en estudiantes de primer semestre de una universidad de la ciudad de Barranquilla. *Psicología Desde El Caribe, 23,* 66-82.

Zvoch, K., & Stevens, J. J. (2003). A multilevel, longitudinal analysis of middle school math and language achievement. *Education Policy Analysis Archives, 11* (20), 1-21.

Zvoch, K., & Stevens, J. J. (2006). Successive Student Cohorts and Longitudinal Growth Models: An Investigation of Elementary School Mathematics *Performance. Education Policy Analysis Archives, 14* (2). Recuperado de epaa.asu.edu/epaa/v14n2/